DOCTOR OF PHILOSOPHY

The Effects of Polysomal mRNA Association and Cap Methylation on Gene Expression in Trypanosoma brucei

Kelner, Anna

Award date:
2014

The Effects of Polysomal mRNA Association and Cap Methylation on Gene Expression in *Trypanosoma brucei*

Anna Kelner

A dissertation submitted in fulfillment of the requirements for

the degree of Doctor of Philosophy

Division of Biological Chemistry and Drug Discovery

College of Life Sciences

2014

CONTENTS

Declaration

I declare that I am the sole author of this thesis and that all references cited have been consulted by me personally. The work, of which this thesis is a record, has been done by myself unless otherwise acknowledged. This work has not been previously submitted for a higher degree.

Signed ……...

Anna Kelner

Date ……

List of Figures

List of Tables

List of Abbreviations

ADP – adenosine diphosphate

ARE – AU-rich element

BSA – bovine serum albumin

BSF – bloodstream form

BP – branch point

bp – base pair

CBC – cap binding complex

CDP – cytidine diphosphate

cKO – conditional null mutant

DEPC – Diethylpyrocarbonate

DIG – digoxigenin

dKO – null mutant

DM – double marker

DMSO – dimethyl sulfoxide

DNA – deoxyribonucleic acid

DPMS – dolicholphosphate mannose synthase

dsRNA – double stranded RNA

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

ER – endoplasmic reticulum

ERAD – ER associated degradation

ESAG – expression site associate gene

FSG – fish skin gelatin

G – guanosine

gDNA – genomic DNA

GDP – guanosine diphosphate

GMP – guanosine monophosphate

GO term – gene ontology terms

GTP – guanosine triphosphate

HAT – Human African Trypanosomiasis

HYG – hygromycin

m7GDP – 7-methylguanosine diphosphate

m7GMP – 7-methylguanosine monophosphate

mRNA – messenger RNA

NaCl – sodium chloride

NaOAc – sodium acetate

NOG – n-octyl-glucoside

nt – nucleotide

O/N – over night

ORF – open reading frame

PABP – poly(A)-binding protein

PAC – puromycin

PAS – polyadenylation sites

PBS – phosphate buffered saline

PCA – principle component analysis

PCF – procyclic form

PCR – polymerase chain reaction

PDI – protein disulphide isomerase

PHLEO – phleomycin

PMSF – phenylmethylsulfonyl fluoride

PRP – paraflagellar rod protein

qRT-PCR – quantitative real-time PCR

RNA – ribonucleic acid

RNAi – RNA interference

RNA Pol II – RNA Polymerase II

rRNA – ribosomal RNA

RT – room temperature

SAM – S-Adenosyl methionine

SAS – slice-acceptor sites

SDS – sodium dodecyl sulphate

SDS-PAGE – sodium dodecyl sulphate polyacrylamide gel electrophoresis

SILAC – stable isotope labelling by amino acids in cell culture

sKO – single knockout

SL RNA – spliced leader RNA

SM – single marker

SSR – strand-switch regions

TBAA – tributylammoniumacetate

TDB – trypanosome dilution buffer

TEAA – triethylammonium acetate

TERT – telomerase reverse transcriptase

tet – tetracycline

TFA – trifluoroacetic acid

TLC – thin layer chromatography

Tris – Tris(hydroxymethyl)aminomethane

UGGT – UDP-Glc:glycoproteinglucosyltransferase

UTR – untranslated region

VSG – variant surface glycoprotein

WT – wild type

Abstract

Contrasting physiological requirements for *T. brucei* survival between procyclic (vector) and bloodstream (mammal) forms necessitate different molecular processes and therefore changes in protein expression. Transcriptional regulation is unusual in *T. brucei* because the arrangement of genes is polycistronic; however, genes which are transcribed together are subsequently cleaved into separate mRNAs by *trans*-splicing and are individually regulated. During the process of *trans*-splicing, a 39-nucleotide splice-leader RNA is added to the 5′ end of mRNA. In this study, gene regulation in trypanosomes will be examined in the context of the 7-methylguanosine cap attached to the 5′ end of the splice-leader. Interestingly, in addition to the capping enzymes identified in other eukaryotes, trypanosomatids have an additional guanylyltransferase and methyltransferase in the form of a bifunctional enzyme (*Tb*CGM1).*Tb*CGM1 was found to be essential in bloodstream form *T. brucei,* althoughthe purpose of this bifunctional capping enzyme remains unclear. Null mutants of a related enzyme, monomeric methyltransferase *Tb*CMT1, did not show an effect on cell viability in culture, however, the enzyme proved to be important for virulence *in vivo*. Complementary to the study of *T. brucei* capping enzymes, we worked to develop a method to allow structural analysis of the 5′mRNA cap by mass spectrometry.

Following pre-mRNA processing, regulation of the mature mRNAs is a tightly controlled cellular process. While multiple stage-specific transcripts have been identified, previous studies using RNA-seq found that the changes in overall transcript level do not necessarily reflect the abundance of the corresponding proteins. We hypothesized that in addition to mRNA stability, mRNA recruitment to ribosomes may play a significant role in the regulation of gene expression in *T. brucei*. To approach this

question, we performed RNA-seqof total, subpolysomal, and polysomal mRNA.This transcriptomic data was then correlated with published proteomic studies to obtain a global picture of the relative translation efficiencies and their relationship to steady-state protein levels between bloodstream and procyclic form *T. brucei.*

Acknowledgements

I would like to sincerely thank Mike for allowing me to work in his lab and for encouraging my research even in difficult times. Despite myriad other obligations, thank you for taking the time to give me advice on a blot or spectra! I would also like to thank Vicky for her support and understanding my baseball references! I also want to thank Lucia for teaching me everything I know about *T. brucei* cell culture and helping me troubleshoot experiments.

I am grateful to be a member of the infamous end-bay, thank you Angela (and her snack drawer) for always helping in a pinch and keeping the lab in high spirits with your laughter. Thanks, Seb, for showing me the effectiveness of organization and punctuality, hopefully I'll learn someday! Anders, it's all your fault.

Thank you to the rest of the MAJF lab, I will show you my gratitude with cake, of course! Thank you, Di and Michele, for answering thousands of my questions and for your patience with me.

As a joint student, I'm lucky enough to have two labs. I would like to thank the VC lab: Dhaval, Laura, Olivia, and Sianadh, your light-hearted mood and friendly smiles always put me in a good mood!

Although they have moved on to greener pastures, I would like to thank Giulia, Nicola and Mick for helping me since the beginning and even now!

I would also like to thank the Horn lab for being such a fun addition to the BCDD, and to Lucy for help with the RNAi experiments and coffee-fuelled support. I was very fortunate to work with a fantastic group of scientists in Matt Berriman's lab at Sanger; I

would like to thank Lia for welcoming me to Cambridge and teaching me how to make RNA-seq libraries and Bernardo for his bioinformatics prowess.

After four years, the strangers who started their PhDs with me have truly become family. Mr. Kelner, thank you for being the world's greatest roomie and debate-loving friend! Lynne, thank you for being my inspiration! Kat, thank you for being and incredible friend and always seeing the positive side of things! …this is just the beginning of our adventures!

I am very grateful to my mom for your unconditional love and support across 4000 miles, thank yougranny for always giving advice and thank you both for being my biggest fans, Я Вас оченьлюблю!! Thank you to my dad, for your endless curiosity and sense of humor!

INTRODUCTION

INTRODUCTION

1.1 Human African Trypanosomiasis

Trypanosoma brucei, the protozoan parasite transmitted by the tsetse fly, causes Human African Trypanosomiasis (HAT) and Nagana in cattle. Following the observations of flagellated parasites (now called trypanosomes) in the blood of frogs and small rodents, in the late nineteenth century, Sir David Bruce discovered the subspecies of trypanosomes (*Trypanosoma brucei brucei*) associated with Nagana which lead to the understanding of the parasite's life cycle and eventual discovery of human infective trypanosomes by R. M. Forde and J.E. Dutton who named them *Trypanosoma gambiense*(Dutton, 1902; Forde, 1902). In 1910, J. W. W. Stevens and H. B. Fantham described a distinctive subspecies of human trypanosome they named *Trypanosoma rhodesiense*(Stephens and Fantham, 1910). It is now known that *Trypanosoma rhodesiense* causes an acute form of HAT in East Africa and *Trypanosoma gambiense* causes the chronic form of the disease in Western and Central Africa (WHO, 2014). Sir David Bruce also pioneered the studies that identified the tsetse fly as the transmission vector (Cox, 2004). HAT, commonly referred to as sleeping sickness, infection manifests itself in two phases: the first stage during which the trypanosomes proliferate in the blood of the host, followed by the second stage when the parasites cross the blood-brain barrier and invade the central nervous system. This second stage is characterized by neurological disorders and is, with rare exception, fatal if left untreated (MacLean et al., 2012; Stuart et al., 2008).

The World Health Organization (WHO) hasled a campaign to eliminate HAT bringing together an alliance of international institutions, organizations and pharmaceutical companies (Simarro et al., 2012). Because sleeping sickness affects poor rural populations that lack access to adequate health facilities, treatment is an even bigger burden (Simarro et al., 2011). The cooperative effort has succeeded in increasing the

control and surveillance of the at-risk populations and as a result, reducing the number of new cases each year by 82%, with only 6,228 new cases reported in 2013 although millions of people are still at risk (Franco et al., 2014; Simarro et al., 2012). Recently, the widespread use of nifurtimox-eflornithine combination therapy (NECT)has been a great advancement in the treatment of HAT with high efficacy, lower toxicity, and shorter course of treatment than previous medication (Alirol et al., 2013; Franco et al., 2012; Priotto et al., 2009; Yun et al., 2010). However, NECT still involves a regimen of invasive procedures that are difficult to administer in regions of poor health infrastructure creating a need for the development of oral drugs that target the second stage of the disease (Barrett *et al.* 2007).

1.2 The life cycle of *T. brucei*

When a Tsetse fly feeds on infected blood from a human or animal, it is infected by stumpy form *T. brucei*trypomastigotes which differentiate into the proliferating procyclic form in the fly midgut(Sbicego et al., 1999). From the midgut, the PCF trypanosomes progress through the salivary gland as epimastigotes and then further differentiate into infective metacyclics(Van Den Abbeele et al., 1999). During a subsequent blood meal, the Tsetse fly transmits the metacyclictrypomastigotes into the haemolymphatic system of the host where they transform into the rapidly proliferating bloodstream slender form, and the cycle begins again (Figure 1.1).

Bloodstream form *T. brucei* are coated by approximately 5 million variant surface glycoprotein (VSG) homodimers that form a dense monolayer and are attached to the cell surface via a GPI anchor (Cross, 1975; Ferguson et al., 1988). These glycoproteins give the parasite a mechanism of immune system evasion.

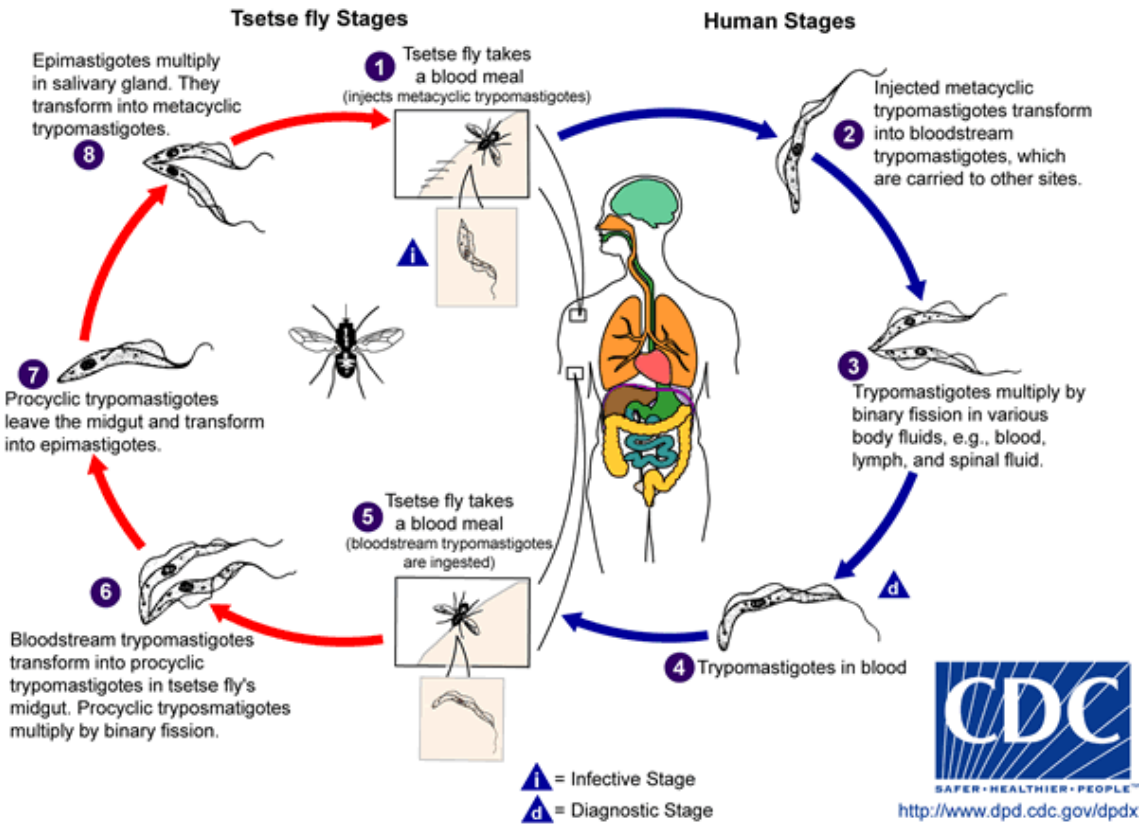


Figure 1.1: *T. brucei* life cycle. Image adopted from www.cdc.gov.

Although VSGs stimulate an immune response in the host which clears most of the infection, a small proportion of the remaining parasites switch their surface coat to a different VSG and repopulate in the bloodstream of the host (Cross, 1979; Horn and Cross, 1997; Pays and Nolan, 1998). This process of VSG switching and the consequent waves of parasitemia are known as antigenic variation (Doyle et al., 1980).

1.3 The *T. brucei* nuclear genome

The nuclear genome of *T. brucei* contains 11 megabase-sized chromosomes, ~5 intermediate-sized chromosomes, and ~100 minichromosomes(Berriman et al., 2005; Melville et al., 2000). Unlike other eukaryotic genomes, the genomes of kinetoplastids (*T. brucei, T. cruzi*and *Leishmania*) share an unusual genomic architecture where adjacent genes form long, non-overlapping transcription units (Figure 1.2). These gene clusters contain between approximately tens to hundreds of protein coding sequences in

the same transcriptional orientation (Berriman et al., 2005). Another difference of trypanosomatid genomic arrangement is that the majority of protein coding genes do not have introns. The exceptions are a poly(A) polymerase and an ATP-dependent DEAD box RNA helicase which each have one intron that is removed by *cis*-splicing (Mair et al., 2000a; Preußer et al., 2014).

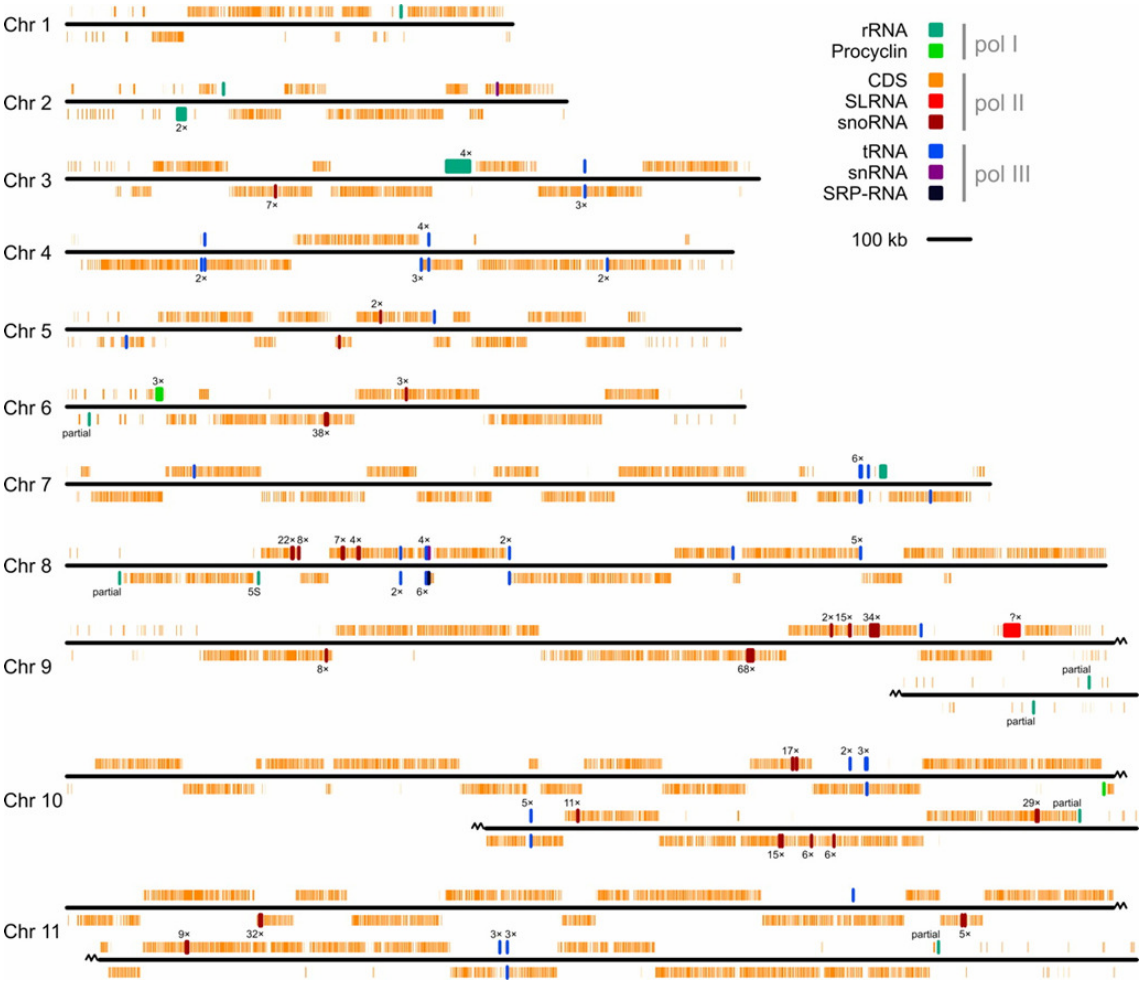


Figure 1.2: Genomic organization in *T. brucei* according to the transcribing polymerase. Bars above the chromosome backbone (black) indicate transcription toward the right; bars below the line indicate transcription toward the left. Similar elements close together in the genomeare indicated by n× (where n is the number of elements). Adopted from (Daniels et al., 2010).

1.4 Polycistronic transcription and *trans*-splicing in *T. brucei*

Transcription is particularly interesting in *T. brucei* because the arrangement of its genes is polycistronic. RNA Polymerase II (RNA Pol II) transcribes protein-coding genes into large polycistrons containing several transcripts, however, the polycistron

does not linger as it is co-transcriptionally processed into individual mRNAs(Johnson et al., 1987; Mottram et al., 1989). These polycistrons are unidirectional and partitioned by strand-switch regions (SSRs) between stretches of genes that are characterized by a stretch of G nucleotides (Figure 1.3). Polycistronic transcription was discovered by *in vitro* run-on transcription assays investigating the mechanism of the transcription of the tubulin gene locus in *T. brucei*(Imboden et al., 1987). It was found that although intergenic regions between tubulin genes were transcribed at the same rate as the genes themselves, transcripts containing intergenic regions could not be detected leading to the conclusion that the transcription of the tubulin gene cluster proceeds as a single contiguous primary transcript which is then rapidly processed into individual mRNAs (Imboden et al., 1987). The processing of the transcription unit occurs co-transcriptionally by *trans*-splicing coupled to cleavage of the 3′ end by the polyadenylation machinery for poly(A) addition (Huang and van der Ploeg, 1991; Ullu et al., 1993).

During *trans*-splicing, a capped 39-nucleotide (nt) spliced leader (SL) mini-exon is added to the 5′ termini of mRNAs. The SL sequence was first discovered when two different VSG transcripts were found with an identical leader sequence at their 5′ ends, which was not evident in their genomic sequence (Boothroyd and Cross, 1982; Parsons et al., 1984; Van der Ploeg et al., 1982). This mini-exon is independently transcribed from a tandem array of 140-nt spliced leader (SL) RNA genes (Figure 1.3) (Gilinger and Bellofatto, 2001; Sather and Agabian, 1985).

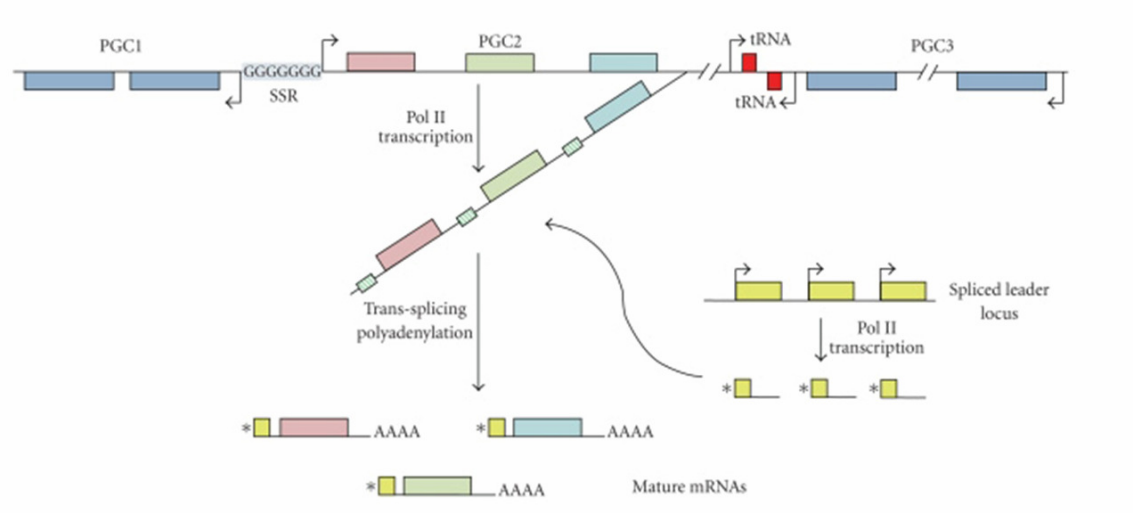


Figure 1.3: mRNA processing in *T. brucei.* Schematic of a hypothetical chromosome with three polycistronic gene clusters (PGC1-3). RNA Polymerase II (RNA Pol II) transcription initiates upstream of the first gene of the PGC (arrows). The stretch of G nucleotides present at strand-switch regions (SSR) is indicated. Transcription of a PGC is processed by *trans*-splicing and polyadenylation to generate the mature mRNAs (shown only for PGC2). In the SL RNA locus, each SL RNA (yellow box) possesses an RNA Pol II promoter region (arrows). The 5' cap is indicated by an asterisk. The pyrimidine-rich regions (shown asgreen-striped boxes in the intergenic regions) are needed for both *trans*-splicing and polyadenylation. Adapted from (Martínez-Calvillo et al., 2010).

The mechanism of *trans*-splicing is a two-step transesterification (Figure 1.4). In the first transesterification step, there is a nucleophilic attack by the 2′OH of the branch point (BP) adenosine from the pre-mRNA onto the 5′ splice site (SS) of the SL RNA. This generates the SL mini-exon intermediate and the SL intron/protein-coding exon 'Y-structure' formation (Murphy et al., 1986). In the second step, the free -OH of the SL mini-exon attaches to the 3′ splice site (SS) of the protein-coding exon. The product of these reactions is the first 39 nucleotides (mini-exon) of the SL RNA precursor and the protein-coding exon are joined together, yielding an mRNA that is ready for polyadenylation (Agabian, 1990; Laird et al., 1987; Sutton and Boothroyd, 1986).

1.5 Structure and function of the *T. brucei* hypermethylated cap4

In addition to acting as a splicing substrate for the excision of an mRNA from the primary transcript, the SL RNA also provides each protein-coding mRNA with the cap structure through *trans*-splicing. For all cells in which it has been investigated, the structure of the mRNA 5′cap was determined to be 7-methylguanosine (Furuichi et al., 1977; Rottman et al., 1974; Shatkin, 1976). Trypanosomatids contain a unique structure at the 5′end of all mRNAs called cap4 (Freistadt et al., 1987, 1988; Perry et al., 1987). The cap4 structure is hypermethylated and consists of a conserved m^7G cap followed by four nucleotides with 2′-*O*-ribose methylations as well as additional base methylations on the first and fourth nucleotides: (m^7G(5′)ppp(5′)$m^6{}_2$AmpAmpCmpm^3Ump) (Figure 1.5) (Bangs et al., 1992; Sutton and Boothroyd, 1988). The remarkable discovery of the trimethyladenosine and the dimethyluridine hypermethylated nucleotides in the cap4 structure were previously unknown in nature (Bangs et al., 1992). Although the complex cap4 structure of the SL RNA is a prerequisite for *trans*-splicing and is therefore essential for the production of mature mRNAs, the mechanism remains unclear (McNally and Agabian, 1992; Ullu and Tschudi, 1991).

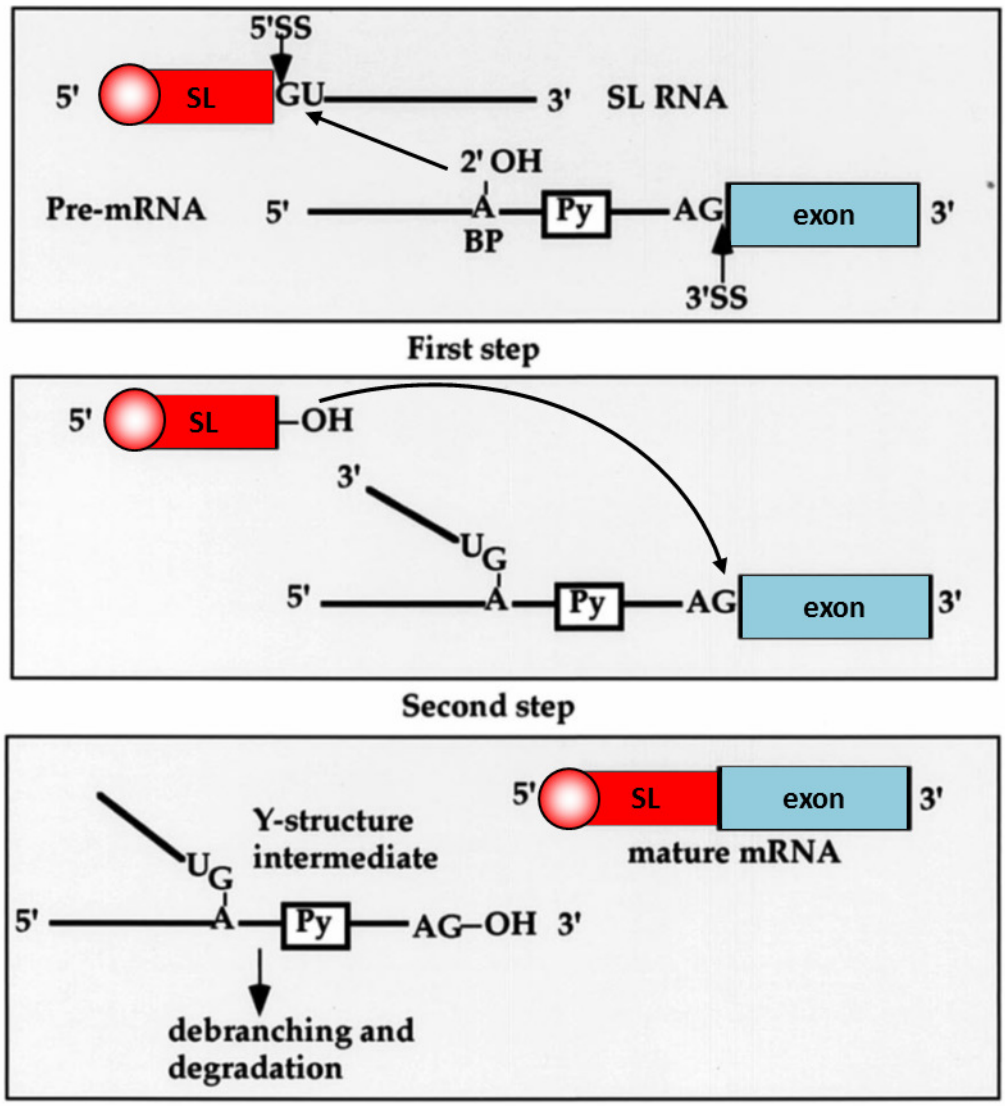


Figure 1.4: Mechanism of *trans*-splicing in *T. brucei*. The SL RNA and a pre-mRNA are pictured, outlining the conserved branch point (BP) adenosine (A), the polypyrimidine tract (Py), a single protein-coding exonof a longer polycistronic transcript (blue box), the SL mini-exon (red box), and the 5'cap ( ). The consensus sequences directly adjacent to the 5′ and 3′ splice sites (5′SS; 3′SS) are also shown. Adapted from (Liang et al., 2003).

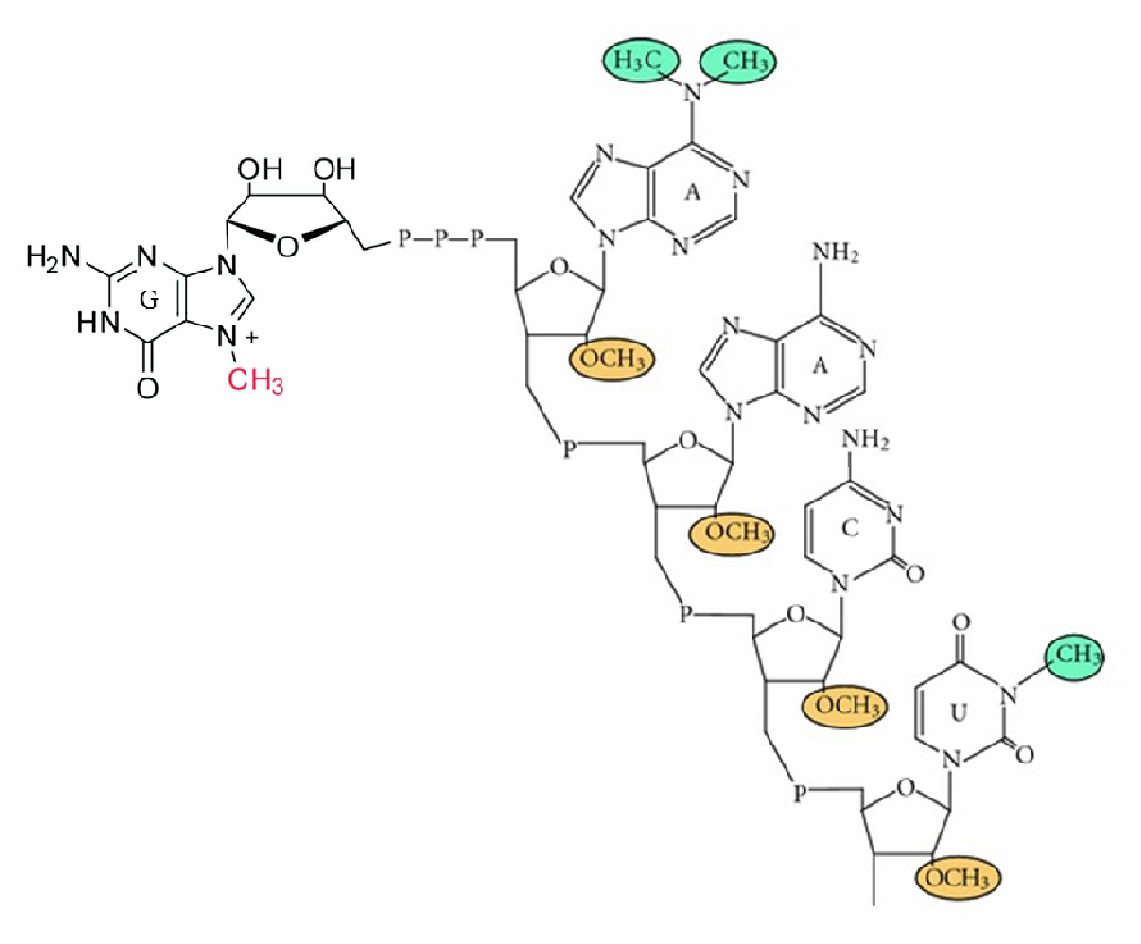


Figure 1.5: Trypanosomatid cap4 structure. Them7G cap is methylated at the N7 position (red). There are also 2-O ribose methylations on the first four nucleotides (orange) as well as base methylations on the first adenosine and uridine (green) of the SL RNA. Adopted from (Bangs et al., 1992; Zinoviev and Shapira, 2012)

Binding an antisense oligonucleotide to nt 7-18 of the SL sequence, and therein preventing modification of the cap4, disabled the use of the SL RNA in *trans*-splicing (Ullu and Tschudi, 1993). That study further supported the importance of the cap4 structure for the fidelity of translation competent mRNA. Modification of the cap structure is not limited to *T. brucei* as the trypanosomatid protozoa contain common machinery for co-transcriptional cap4 formation (Mair et al., 2000b; Ullu and Tschudi, 1995).

1.6 *T. brucei* capping enzymes

The importance of the modification at the 5′ end of *T. brucei* SL RNA lead to the characterization of the enzymes involved in this process.Cap formation occurs in three enzymatic reactions. First, the 5′ triphosphate end is hydrolysed by a triphosphatase to produce diphosphate–RNA. Next, the guanylyltransferase catalyses the addition of GMP to the diphosphate–RNA to produce the guanosine cap. Finally, the cap is acted upon by a guanine-7-methyltransferase, catalysing the methylation of the guanosine cap at the N-7 position. The capping system in metazoans and plants consists of a bifunctional triphosphatase- guanylyltransferase and a separate methyltransferase (Figure 1.6) (Cowling, 2010; Shuman, 2001, 2002). In *T.brucei*, the first characterized enzyme of the capping machinery was the mono-functional guanylyltransferase (*Tb*CE1), which transfers GMP to the diphosphate end of RNA. There was no noticeable effect on PCF *T.brucei* cell viability when this activity was down-regulated by RNAi(Ruan et al., 2007; Silva et al., 1998; Takagi et al., 2007). The T. brucei RNA triphosphatase (*Tb*CET1) catalyzes the removal of the γ-phosphate from triphosphate-terminated RNA and is a separately encoded enzyme that is structurally and mechanistically closer to the fungal and RNA triphosphatase than to metazoan RNA triphosphatases(Ho and Shuman, 2001). A (guanine N-7) methyltransferase (*Tb*CMT1) has also been characterized (Hall and Ho, 2006a). Sequence alignment between *Tb*CMT1 and the corresponding enzymes from humans, fungus, and microsporidia showed several conserved residues that are essential for methyltransferase activity (Hall and Ho, 2006a). While *Tb*CMT1 can methylate the cap analogues GpppA and GpppG, GTP is not a suitable methyl acceptor, suggesting *Tb*CMT1 can discriminate free nucleotides to only methylate RNA. On the other hand, its ability to methylate cap analogues also suggests the RNA component beyond the first nucleotide is not strictly

essential for catalysis (Hall and Ho, 2006a). Knockdown of *Tb*CMT1 by RNAidid not show a growth defect in PCF *T. brucei* (Takagi et al., 2007).

Interestingly, in addition to the three capping enzymes described, trypanosomatids have an additional guanylyltransferase-methyltransferase in the form of a bifunctional enzyme (*Tb*CGM1) (Figure 1.6, panel B) (Hall and Ho, 2006a; Takagi et al., 2007).In the characterization of this unusual capping enzyme, it has been shown that the N-terminal domain contains the six signature guanylyltransferase motifs that are conserved in the superfamily of capping enzymes and form the active site for GMP binding and nucleotide transfer (Ruan et al., 2007; Takagi et al., 2007). The C-terminus of *Tb*CGM1 contains an S-Adenosyl methionine (SAM) binding motif characteristic of methyltransferases as well as several important residues that are involved in the binding of the GpppA cap (Hall and Ho, 2006a; Takagi et al., 2007). The concerted mechanisms of this enzyme form the 5′ terminal (cap0) structure on SL RNA. *Tb*CGM1 depletion by RNAisupported its essentiality in PCF *T. brucei* (Ruan et al., 2007; Takagi et al., 2007). Silencing of *Tb*CGM1 by RNAi also causes defects in cap4 formation seen by the accumulation of hypomethylated SL RNA and therein reduced utilization of SL RNA in *trans*-splicing (Ruan et al., 2007; Takagi et al., 2007). Cap0 formation is important as it is a pre-requisite for subsequent methylations the SL RNA to form cap4 because the cap2 methyltransferase requires the structure m7GpppA for binding and catalysis (Hall and Ho, 2006b; Mittra et al., 2008; Takagi et al., 2007). These studies have laid the groundwork for understanding the *in vivo* functions of the capping machinery in *T. brucei*; however, the purpose of the multiplicity of trypanosomatid capping enzymes remains unclear.

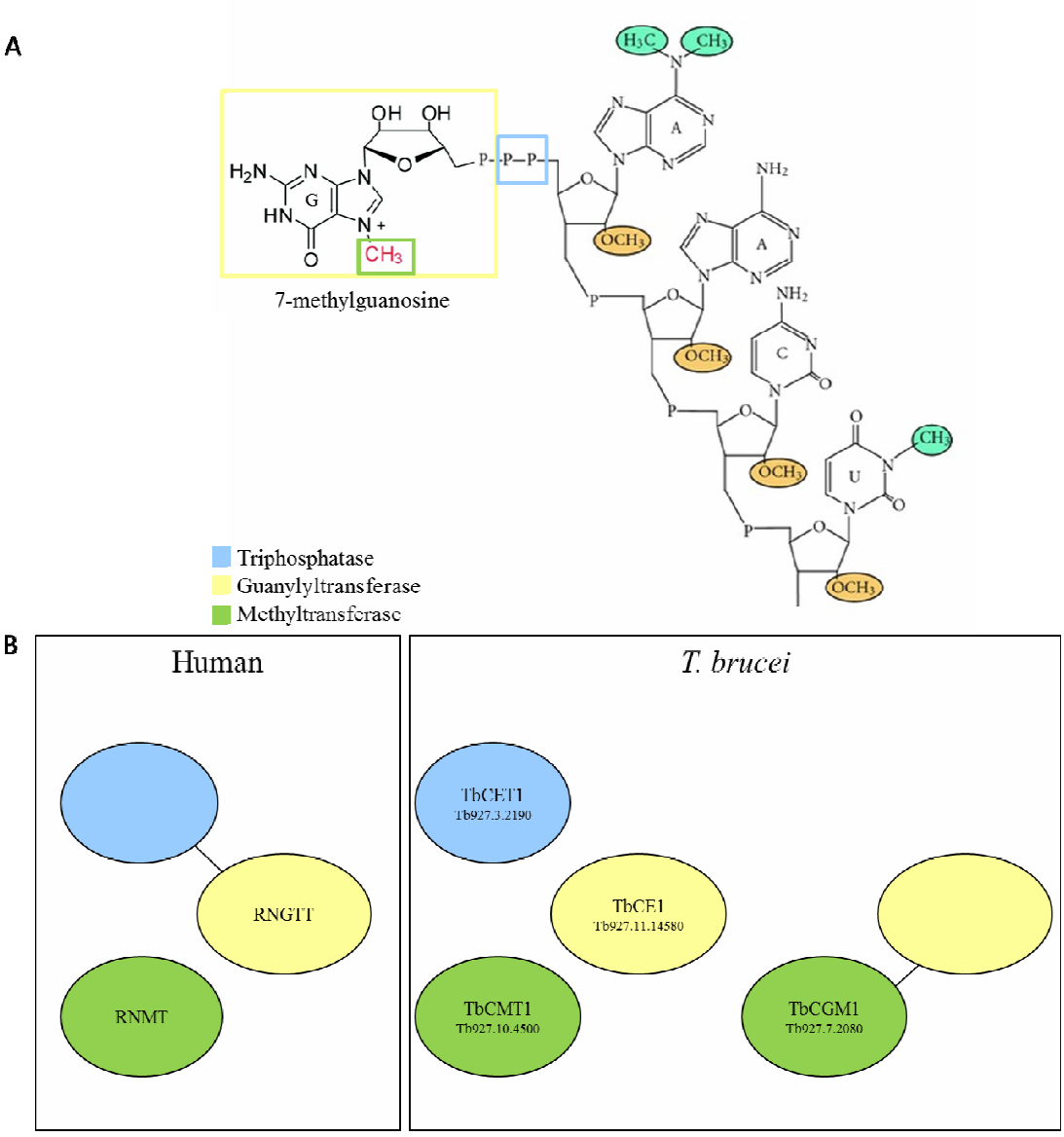


Figure 1.6: The mRNA capping apparatus differs between humans and *T. brucei.* (A) Structure of the 7-methylguanosine cap (cap0) and first four transcribed nucleotides, collectively known as cap4. The action of each capping enzyme is depicted by ovals of the corresponding color. (B) Humans have a triphosphatase (blue) fused to the guanylyltransferase (yellow) and a methyltransferase (green). *T. brucei* have the three enzymes encoded separately plus an additional guanylyltransferase-methyltransferase bifunctional enzyme.

1.7 Post-transcriptional regulation of gene expression in *T. brucei*

Although in *T. brucei*genes are arranged in tandem arrays and transcribed as polycistrons, they are generally unrelated in function or developmental regulation. Modulation of mRNA half-life occurs, at least in part, through RNA-binding proteins which interact with motifs located in the 3´ untranslated region (UTR) (Clayton, 2002, 2013). For example, AU-rich elements (AREs) are motifs which stimulate 3´ to 5´exonuclease activity. Additionally, mRNA turnover is regulated through the activity of a decapping enzyme that removes 7-methylguanosine diphosphate (m7GDP) from the end of mRNA, as well as by a decapping scavenger activity that interacts with the exosome-mediated decay machinery (Irmer and Clayton, 2001; Milone et al., 2002) . These degradation pathways are counterbalanced with binding sites for poly(A)-binding proteins (PABPs) which prevent mRNA 5´decapping(Hotchkiss et al., 1999). Together, these processes allow for dynamic adaptation to variable environments via sub-regions in the 3´UTRs that have different effects on mRNA stability throughout the life cycle (Milone et al., 2002). The first life cycle specific regulatory sequences identified in *T. brucei* were in the 3´ UTR of *VSG* mRNA (Berberof et al., 1995). In BSF these sequences stimulated expression by stabilizing the mRNA and conversely, in PCF, *VSG*translation was shut down through instability and rapid mRNA degradation. Since then, similar regulatory motifs have been characterized for several genes (Clayton, 2002). The extensive regulatory capacity of RNA-binding proteins is seen in the study of RBP6, where over-expression allowed the recapitulation of the *T. brucei* life cycle from the procyclic form trypomastigotes in the tsetse fly midgut to the host infective metacyclictrypomastigotes(Kolev et al., 2012).

In the nucleus, mature mRNAs that have evaded the degradation machinery are recognized by the cap binding complex (CBC) via their cap4 structure. Trypanosomatids have an unusual CBC consisting of five subunits (compared to the

heterodimeric CBC in other eukaryotes) that is essential for cell viability. *T. brucei* CBC binds cap4-containing RNA with significantly greater affinity than m^7G-capped RNA, emphasising its uniqueness (Li and Tschudi, 2005). When the CBC dissociates from the cap4-RNA, it is subsequently transported into the cytoplasm and bound by the cap binding translation initiation factor eIF4E which is a component of the heterotrimeric translation initiation complex alongside an RNA helicase, eIF4A, and the scaffolding protein eIF4G (Freire et al., 2011; Pereira et al., 2013). The *T. brucei* family of eIF4E cap4 binding proteins has six homologs whose interaction with eIF4G mediates the interaction with the newly characterized hypothetical protein eIF4G5-interacting protein (*Tb*G5-IP) (Freire et al., 2014). This complex facilitates the recruitment of the 40S ribosomal subunit to the mRNA and therein translation. The existence of multiple initiation factor homologs might suggest the complex serves as an mRNA selector to create a co-regulated response to differentiation. These processes of transcription, post-transcriptional regulation and translation are consistent between bloodstream and procyclic form *T. brucei* although in order to survive the contrasting environments between their host and vector, the parasites must alter their gene expression profiles.

1.8 Differential gene expression throughout the life cycle of *T. brucei*

Continuous advances in analytical tools in transcriptomics and proteomics fuel our ability to understand interactions and mechanisms unattainable in the past. Trypanosome progression from mammalian blood to the flymidgut requires rapid changes in protein expression to replace its VSG surface coat with procyclin, and alter its metabolism. Further, the parasite must then differentiate into its metacyclic form in the salivary glands to restart the cycle. Recent studies using RNA-seq have greatly improved our understanding of the *T. brucei* transcriptional landscape. These studies have found approximately a thousand new transcripts, multiple non-coding RNAs, and

facilitated the correction of numerous annotations across the *T. brucei* genome. The three research groups who performed these studies observed 5.6-32% of all mRNAs were regulated,by at least 2-fold,between BSFand PCF suggests *T. brucei* controls gene expression post-transcriptionally (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010). Using RNA-seq it was possible to precisely map previously undefined UTR boundaries for genes. The length of a gene's 5′ UTRs was found to be, on average, shorter than its 3′UTR, presumably to accommodate the regulatory motifs found in the 3′UTR. Interestingly, multiple splice-acceptor sites (SAS) and polyadenylation sites (PAS) were discovered per gene, revealing a broad range of alternative SAS and PAS that have not been characterized (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010). This result implies the extent of gene regulation by different splice variants has been underestimated and could lead to interesting research in the future. Further, for 321 genes, all of the SAS mapped to an initiator ATG site that is downstream of the previously annotated coding sequence start-site (Siegel et al., 2010). Additionally, RNA-seq enabled the discovery of over a thousand transcripts that were not annotated in the *T. brucei* genome which contain open reading frames (ORFs) and 50 of these transcripts contain orthologs in *Leishmania* and *Trypanosoma cruzi*. Of these novel transcripts, 111 are developmentally regulated between BSF and PCF trypanosomes and 19 have been detected at the protein level by mass spectrometry (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010). These findings are a transcriptomic gold mine of potential regulators of the *T. brucei* life cycle awaiting discovery.

1.9 Translational regulation of gene expression in *T. brucei*

Investigation of translational control on a global scale has been largely overlooked in *T. brucei*. Nevertheless, there are a few examples of polysome profile analysis used to explore differential translation in *T. brucei*, and recently, the ribosomal profiling method has been implemented for the study of *T. brucei* gene expression regulation

(Brecht and Parsons, 1998; Capewell et al., 2013; Vasquez et al., 2014). The quantification of translated RNA transcripts is achieved by isolating the transcripts from the polyribosome fractions of a sucrose gradient (Dickson and Brown, 1998). Although ribosomes could be associated with the 5′ UTR or stalled, polysome associated transcripts are considered more likely to be actively translated (Gross et al., 2003; Ingolia et al., 2009; Sagliocco et al., 1993). The ribosome profiling method involves isolating ribosome-protected RNA fragments combined with RNA-seq to map the ribosome 'footprints'. Protein expression is then estimated from the density of the ribosome footprints (Ingolia et al., 2009). In *T. brucei*, ribosomal profiling showed a large variation in translational efficiency both within an individual life cycle stage, and comparatively between bloodstream and procyclic form cells. This suggests the translation of mRNA to protein is an important regulatory step in *T. brucei,* consistent with observations in other eukaryotes (Ingolia et al., 2009, 2011; Vasquez et al., 2014).

1.10 Comparative transcriptomics and proteomics in *T. brucei*

There are many genes for which the reported life cycle regulation of mRNA does not correlate with the observed protein levels: RAB18, transketolase, and CAP5.5 to name a few (Hertz-Fowler et al., 2001; Jeffries et al., 2002; Stoffel et al., 2011). Unlike the small number of life stage-regulated transcript changes (5.6-32%) discussed above, changes in the protein levels vary drastically between life stages. However, the relationship between transcript and protein abundance is not yet well understood (Urbaniak et al., 2012, 2013). The correlation between stable isotope labelling by amino acids in cell culture (SILAC) comparative proteomic data and microarray transcriptomics studies in *T. brucei* was examined and the mRNA abundance from two different studies (Jensen et al., R=0.86, Queiroz et al., R=0.83) correlated well with protein (Jensen et al., 2009; Queiroz et al., 2009; Urbaniak et al., 2012). Strong correlation between mRNA and protein levels combined with RNA-seq and

translational profiling have made important contributions to the understanding of how trypanosomes adapt throughout their life cycle.

1.11 Aims of the thesis

Aim 1:To explore the essentiality and function of the bifunctional capping enzyme and monomeric methyltransferase. This will be accomplished by:

a) Generation of conditional null mutants of the monomeric methyltransferase (*Tb*CMT1) and the bifunctional methyltransferase/guanylyltransferase (*Tb*CGM1).
b) Investigating the effects of loss of *Tb*CMT1 function on *in vivo* cell viability, the *T. brucei* transcriptome and translation.

Aim 2: To investigate the possibility that some of the mRNA 5´ cap nucleosides in *T. brucei* are structurally different from conventional 7-methylguanosine. To achieve this, a method was developed to isolate the cap from endogenous mRNA and analyse the resulting structure by mass spectrometry. The possibility of novel mRNA cap structure(s) was suggested by the existence of an additional guanylyltransferase/methyltransferase (*Tb*CGM1) found in kinetoplastids but not other eukaryotes. Specifically, a method will be developed to isolate endogenous mRNA capping nucleosides from trypanosome mRNA preparations and identify the resulting structure(s)using LC-MS/MS as the primary analytical tool. The methodology will be developed using commercially available methyl cap standards and, subsequently, human (HEK293 cell) mRNA.

Aim 3: To investigatemRNA recruitment to ribosomes as a component of gene expression regulation in trypanosomes using RNA-seq of total polyA+, subpolysomal, and polyribosomal mRNA purified from bloodstream and procyclic form *T. brucei,* followed by bioinformatic analysis comparing transcript and proteomic profiles.

Categorized genes that are over/under-represented on polysomes relative to total mRNA will be used to ask the following questions:

a) What biological processes and pathways are genes that are over/under represented on polysomes associated with?
b) Can we perturb these pathways and observe transcripts moving into/out of polysomes?
c) Do polysomal mRNA ratios between BSF and PCF trypanosomes have a better correlation with observed protein levels than total mRNA?

MATERIALS AND METHODS

MATERIALS AND METHODS

2.1 *Trypanosoma brucei* cell culture growth conditions

Bloodstream form *T. brucei* cell line and media

T. brucei bloodstream form cells (strain 427, VSG variant MIT-aT 1.2) were cultured at 37°C with 5% CO_2 in cell culture flasks with filter lids (Greiner). Cells were grown to a maximum density of $3x10^6$cells/ml in HMI-9t medium. HMI-9t contains variations on the HMI-9 medium described in(Hirumi and Hirumi, 1994): thioglycerol(Sigma) was used instead of β-mercaptoethanol, and GlutaMAX (Gibco) was used instead of L-glutamine in both cases because of increased stability. All null and conditional null mutants were generated from the single marker (SM) cell line, which expresses a T7 polymerase and a tetracycline repressor protein (TetR) under neomycin (G418) selection (Wirtz et al., 1999). Table 2.1 lists the concentration of antibiotics used to maintain selection.

T. brucei procyclic form transgenic cell line 29.13.6 was cultured at 28°C in Becton Dickinson culture flasks. Cells were grown to a maximum density of $4x10^7$ cells/ml in SDM-79 medium (Brun and Schonenberger, 1979) supplemented with 15% fetal bovine serum (FBS) (PAA), GlutaMAX (Gibco), and 15µg/ml hemin (Sigma). Known as the double marker (DM) cell line, these cells were used for polysomal profiling and RNA-seq. Similar to SM cell line, the DM cell line was modified to contain a T7 polymerase and a tetracycline repressor protein (TetR) under G418 and hygromycin (HYG) selection(Wirtz et al., 1999). Table 2.1 lists the concentration of antibiotics used to maintain selection.

Table 2.1: Antibiotics used in *T. brucei* cell culture and genetic manipulation

Antibiotic	Resistance Gene	Concentration (µg/ml) Bloodstream form	Procyclic form
Neomycin (G418)	Neomycin phosphotransferase	4	15
Puromycin	Puromycin acetyltransferase	4	50
Phleomycin	Phleomycin resistance protein	0.1	
Hygromycin	Hygromycin phosphotransferase	1	
Tetracycline		0.5	

Tetracycline (Calbiochem) was prepared fresh every seven days at 10 mg/ml in 70% ethanol and added to the cell culture every two days to maintain expression from the tetracycline inducible promoters. For growth curves of conditional null mutants, tetracycline was removed from the culture by washing the cells three times in medium without the drug and then inoculating cultures in the presence or absence of tet. For RNAi experiments, cells were induced with 1 µg/ml tetracycline every 24 h for effective knockdown.

Both bloodstream form and procyclic form *T. brucei* were counted using the CASY Cell Counter + Analyser system. To stay within the linear range of the cell counter, procyclic form cells were diluted 1:10 in PBS before being diluted in the CASY-ton solution for measurement. The system displays the average of three measurements of the culture's cell density as well as average cell volume.

Generating stabilates

Bloodstream form *T. brucei* were grown to $2x10^6$ cells/ml, centrifuged (800 g, 10 min, 4°C in 4K15 Sigma centrifuge). The pellet was resuspended in 1 ml 'freezing media' (HMI-9t medium without antibiotics and with 10% glycerol) per every 3 ml of culture. The cells were then aliquoted 1 ml per Cryovial (Nalgene). Procyclic form *T. brucei* were grown to $2x10^7$ cells/ml, centrifuged (600 g, 10 min 4°C). The pellet was then resuspended in 1 ml 'freezing media' (SDM-79 with 10% glycerol) per every 5 ml of initial culture. The cells were then aliquoted1 ml per Cryovial (Nalgene). For both cell

lines, the Cryovials were either frozen overnight in a freezing chamber containing isopropanol (Nalgene) or in a Revco Benchtop Ultra-low temperature -80ºC freezer and then stored in a K-series cryostorage unit (Jencons).

2.2 Protocols for generation of null and conditional null mutants

Gene replacement constructs

The open reading frames and flanking regions of genes were obtained using *T. brucei* database (www.tritrpydb.org) using the gene accession numbers: monomeric methyltransferase *TbCMT1* – Tb927.10.4500, guanylyltransferase-methyltransferase bifunctional enzyme *TbCGM1* – Tb927.7.2080, and monomeric guanylyltransferase *TbCE1* – Tb927.11.14580. Plasmid construct sequences were assembled *in silico*using CLC Main Workbench v6 software (CLC Bio). The primers used to amplify UTRs for the gene replacement construct and the ORF for overexpression vectors are listed in Table 2.2. Primers were designed to contain ~20 bp of complementary sequence, have a guanine or cytosine at their 5′ and 3′ ends, a melting temperature (T_m) around 60°C, and less than 2°C difference in T_m between a pair of primers. The primers were designed using the NCBI-PrimerBLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) which uses the program Primer3 to design the PCR primers and then uses the BLAST algorithm to screen the primers against the *T. brucei* genome in order to avoid primer pairs that can create non-specific amplifications(Ye et al., 2012).All of the primers were synthesized by Eurofins Genomics.

Table 2.2: List of primers for gene replacement constructs

Gene	Name	Sequence	Function
TbCGM1 Tb927.7.2080	5' UTR – F	ATA AGT ATG CGG CCG CCA AAT CTA ACA GAA ACA CCC CAA G	5' UTR – Forward
	5' UTR – R	GTT TAA ACT TAC GGA CCG TCA AGC TTT CCC CTC CGA GTG ACT ATT CC	5' UTR – Reverse
	3´ UTR – F	GAC GGT CCG TAA GTT TAA ACG GAT CCG GCG TGT CGG CCA GCA CTG AAA AC	3´ UTR – Forward
	3´ UTR – R	ATA AGT AAG CGG CCG CCG ACC GCG ACC GCC ACC TC	3´ UTR – Reverse
	ORF – F	ATA AGT ATC ATA TGG AAT CGA CGA GTT CAC TTT CGG	ORF – Forwardfor pLEW 100v5
	ORF – R	ATA AGT AAC TCG AGC TAA TCT TGA CGA ACA AGA AGG	ORF – Reverse for pLEW 100v5
TbCMT1 Tb927.10.4500	5' UTR – F	ATA AGT ATG CGG CCG CGC ACT CGC AGC GCT ATC CAG TTA TCC	5' UTR – Forward
	5' UTR – R	GTT TAA ACT TAC GGA CCG TCA AGC TTT AAG GTT ACG CTT TCA CCC CTT	5' UTR – Reverse
	3´ UTR – F	GAC GGT CCG TAA GTT TAA ACG GAT CCG GAG TAC TTA TCT CCC CGT TTT C	3´ UTR – Forward
	3´ UTR – R	ATA AGT AAG CGG CCG CGC TGG CAT ACA GGT GAC TGG CTT C	3´ UTR – Reverse
	ORF – F	ATA AGT ATC ATA TGG AGA GCC TAC GGA CTG CAG C	ORF – Forwardfor pLEW 100
	ORF – R	ATA AGT AAC TCG AGC TGC TGG CTT TCC GGA AGC ACA AC	ORF – Reversefor pLEW 100

Table 2.3: List of plasmids

Plasmid	Use	Source
pGEM5zF	gene replacement constructs	PromegapLEW
100-3xMYC	tetracycline inducible expression of ectopic copy gene with 3 C-terminal MYC tags	Wirtz, et al., 1999
pLEW100v5-3xHA	tetracycline inducible expression of ectopic copy gene (~10 fold greater than pLEW 100) with 3 C-terminal HA tags	Wirtz, et al., 1999 modified by L. Güther

Polymerase chain reactions (PCRs) for amplification of UTRs, ORFs, and Southern probe templates were performed using KOD Hot Start DNA polymerase (Novagen) in a PTC-225 Peltier Thermal Cycler (MJ Research) with the following reaction conditions and programs.

Table 2.4: PCR reaction components and amplification program

Reaction Mix (per 50 µl)
5 µl of 10x KOD Hot PCR Buffer
5 µl dNTPs (0.2mM each)
2 µl $MgSO_4$(1mM)
template (50 ng for plasmid, 100 ng for gDNA)
1.5 µl 5' primer (15 pmol)
1.5µl 3´ primer (15 pmol)
1 µl KOD Hot Start DNA polymerase (1U)
up to 50 µl with nuclease-free H_2O

Amplification Program			
	#of Cycles	T (ºC)	Time
	1	95	2 min
x30	1	95	15 sec
	1	55(T_m-5ºC)	30 sec
	1	72	20sec/kbp
	1	72	2 min
		4	∞

When joining the UTRs in the gene replacement construct, an alternate PCR program was used. The two UTRs were fist annealed via the pre-determined complementary sequence (in the absence of primers) followed by amplification of the combined fragment by the addition of KOD Hot Start DNA polymerase and the 5´ UTR – F and 3´ UTR – R primers (Figure 3.1, Table 2.5).

Table 2.5: Annealing PCR reaction for joining UTRs

Reaction Mix (per 50 µl)
5 µl of 10x KOD Hot PCR Buffer
5 µl dNTPs (0.2mM each)
2 µl $MgSO_4$(1mM)
50 ng each UTR fragment
1 µl KOD Hot Start DNA polymerase (1U)
up to 50 µl with nuclease-free H_2O
after linking PCR add:
1.5 µl 5' UTR – F primer (15 pmol)
1.5µl 3´ UTR – R primer (15 pmol)

Amplification Program			
	#of Cycles	T (ºC)	Time
	1	95	30 sec
x5	1	95	30 sec
	1	50	45 sec
	1	68	1 min
	1	95	30 sec
x25	1	95	30 sec
	1	60	45 sec
	1	68	1 min
	1	68	10 min

PCR reaction products were separated by agarose gel electrophoresis. DNA was analysed on 1 % w/v agarose (Electran agarose, VWR) gels in 1x TAE buffer (40 mMTris-acetate, 1 mM EDTA) containing 0.4 µg/ml ethidium bromide. For Southern blotting, 0.8% gels were used under the same conditions. The DNA samples were diluted with 6x Blue/Orange Loading Dye (Promega) and the 1 kb DNA ladder (Promega) was used to estimate the size of DNA fragments. The agarose gels were run at 80 V in 1x TAE buffer using BioRad Mini-Sub Cell GT tanks connected to Power Pac 300 power packs. Gels were imaged on a UV Transilluminator and captured on aUGenius gel documentation system (Syngene). Following confirmation the PCR products were the correct size, the remaining PCR reaction mixture was purified using the QIAquickPCR Purification Kit (Qiagen) per manufacturer's protocol. DNA concentration was estimated by A_{260} and purity by A_{260}/A_{280} and A_{260}/A_{230} on the nanodrop 2000c spectrophotometer (Thermo).

All restriction enzymes were purchased from NEB. Restriction digests were performed in the recommended buffer O/N in a 37°C stationary incubator for cloning or plasmid linearization prior to *T. brucei* electroporation. For analytical digests to screen several

selected plasmid clones, restriction digest was carried out for 2 h in a 37°C water bath. Restriction digest products were separated by agarose gel electrophoresis as described above and the correct size DNA bands were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen) per manufacturer's protocol. After endonuclease digestion, plasmids were dephosphorylated to reduce probability of re-ligation. The plasmids were incubated at 37°C for 30 min with Antarctic Phosphatase (NEB) followed by heat inactivation at 65°C for 20 min.The concentrations of digested plasmid vector and insert were estimated by measuring the A_{260}on the spectrophotometer. Three different ratios of insert to vector were tried for each ligation, 1:1, 1:3 and 1:7 (vector:insert). A ligation reaction without insert was set up as a control for vector re-ligation. The ligation reactions were set up in 10 µl volumes using the T4 DNA ligase and associated 10x ligation buffer (Promega).The reactions were incubated at RT for 2 h and 1µl of each was transformed into Silverselect competent cells (Invitrogen). Competent cells were transformed by incubating the plasmid with 50 µl competent cells on ice for 30 min, heat shocking the cells at 42ºC for 45 sec, allowing the cells to recover from the heat shock on ice for 2 min, adding 250 µl SOC media to allow the cells an outgrowth period (without selection antibiotic) of 1 h in a shaking 37ºC incubator, plating the transformed cells onto LB+ampicillin (amp) agar plates and incubating the plates at 37ºC O/N. Once colonies were selected and grown O/N in 5 ml LB+amp medium, the plasmid DNA was purified using a Miniprep Kit (Qiagen)and its sequence verified by endonuclease digestion and sequencing (University of Dundee, College of Life Sciences, DNA sequencing service). For stabilates of *E. coli*, a well grown (16h) culture was supplemented with 10% glycerol (AnalaR, VWR) and stored at -80°C. Stabilated*E. coli* cells were revived by inoculation into LB+amp media and grown overnight at 37°C. All media were obtained from the Media Kitchen service in the College of Life Sciences, University of Dundee.

Completed constructs (10 µg) were linearized by digestion with the *NotI* (NEB) restriction enzyme as described above and purified by ethanol precipitation. For each volume of digested plasmid DNA, 3M sodium acetate pH 5 was added at 1/10 of the volume, followed by 2 times the volume cold 100% ethanol. The solutions were mixed and incubated at -80°C for at least 2 h. After incubation, the precipitated DNA was pelleted by centrifugation (30 min, 16,000 x *g*, 4°C). The supernatant was carefully decanted to not disturb the pellet which was then washed with 1 ml cold 70% ethanol. After decanting the wash, the ethanol was left to air-dry for 10 min and then the pellet was resuspended in sterile water at a concentration of 200 ng/µl.

Bloodstream form *T. brucei* electroporation

Gene replacement constructs were prepared for electroporation by digestion with the *NotI* restriction enzyme to release the resistance cassette and UTRs from the pGEM vector backbone. Bloodstream form cells grown to mid log phase density ($2x10^6$ cells/ml), were counted and $1x10^7$ cells/cuvette were centrifuged (800 x *g*, 10 min, RT) and resuspended in 100 µl/cuvette of either in Amaxa Human T-Cell Nucleofector solution containing supplement (Lonza) or Cytomix solution (2mM EGTA p.H. 7.6, 120mMKCl, 0.15mM$CaCl_2$, 10mM K_2HPO_4/KH_2PO_4 pH 7.6, 25mM HEPES pH 7.6, 5mM$MgCl_2 \cdot 6H_2O$, 0.5%Glucose (Dextrose), 100 µg/mlBSA, 1mMHypoxanthine. Plasmid DNA (1µg) was combined with the 100 µl of cell suspension into each of7 cuvettes (Amaxa certified, included in the Nucleofector kit, or BioRad) and 1 cuvette with water in place of DNA as a negative control and measure of the amount of time necessary for antibiotic selection. The cells were electroporated using an AmaxaNucleofector II, program X-001 (Burkard *et al*., 2007). After the electroporation, the cells from each cuvette were recovered in a flask with 12 ml of HMI-9t without the

selection antibioticfor 14 h at 37°C before 12 ml of medium containing twice the concentration of selection antibiotic (puromycin or hygromycin) was added to each flask.

When the pLew100 or pLEW100v5 constructs were prepared for electroporation, plasmids were linearized with NotI restriction enzyme and purified by ethanol precipitation. For electroporation, 2 cuvettes were transfected with the construct and 1 negative control cuvette with water. HMI-9t media containing phleomycin was added to the flask after the usual recovery time at 37°C. Antibiotic selection was deemed complete when cells that were electroporated without DNA were completely dead. Resistant cells were plated into 96-well plates (Grenier) at a concentration of 1 cell/ml to ensure the resistant cells were a clonal population. Cells from five of the wells were chosen to establish in culture and their genotype was analysed by Southern blot.

2.3Protocols for characterization of null and conditional null mutants

Isolation of bloodstream form *T. brucei* genomic DNA

Log phase density bloodstream form *T. brucei*($1x10^8$cells) werecentrifuged (800 x*g*, 10 min, 4°C). The cells were then lysed by resuspending in 1 ml DNAzol (Helena Biosciences) and incubated for 30 min at RT. To precipitate the gDNA, 0.5 ml of absolute ethanol was added, inverted gently until the mixture was homogeneous, and then centrifuged (5000x*g*, 10 min, RT) in a microcentrifuge, Eppendorf. The supernatant was discarded and the gDNApellet was washed with 70% ethanol (5000x*g*, 2 min, RT). The wash was decanted and the pelleted gDNA was left to air-dry for 5 min before being re-dissolved in 50-100 µl of nuclease-free water at 50°C. The concentration of the purified genomic DNA was estimated by A_{260} and purity by A_{260}/A_{280} and A_{260}/A_{230} on the nanodrop 2000c spectrophotometer (Thermo).

Southern blot analysis

Approximately 5 µg of genomic DNA were digested at 37°C for 16 h with chosen restriction endonucleases in their appropriate 10x buffer and the fragmented gDNA was run on a 0.8% agarose gel (in a longer gel tray to increase resolution of genomic fragments) with ethidium bromide and imaged as described above for PCR products. The gel was then prepared for transfer by washing it for 10 min in 0.2 M HCl, then 15 min in 0.5M NaOH, 1.5M NaCl and finally for 20 min in 1M Tris-HCl pH 7.5, 1.5M NaCl. All of the washes were carried out at room temperature while shaking. The gDNA was then transferred on to a positively charged nylon membrane (Roche) by capillary action (Figure 2.1). The transfer was performed for a minimum of 2 hr (usually overnight) in 10x sodium chloride, sodium citrate (SSC) buffer (3M NaCl, 300 mM $Na_3C_6H_5O_7$, pH 7) provided by the Media Service Facility, College of Life Sciences, University of Dundee. After the transfer, the DNA was cross-linked to the membrane using 1200 UV counts on a CL-1000 (UVP) UV cross-linker.

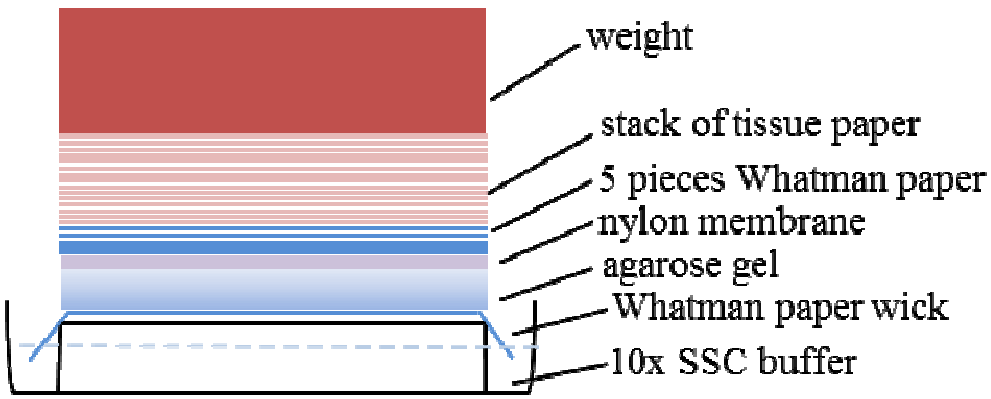


Figure 2.1: Southern blot transfer set up of fragmented gDNA from the agarose gel to the nylon membrane.

The DNA probe for the Southern blot was synthesized using the PCR DIG Probe Synthesis Kit (Roche). A control PCR without the DIG labelled dNTPs was also performed to ensure the PCR synthesis was successful. Both PCR reactions were then analysed by gel electrophoresis to verify the size increase of the labelled probe and approximate the amount synthesized. After crosslinking, the membrane was placed in a roller bottle (Techne) and incubated with 10 ml of the pre-hybridization EasyHyb solution (Roche) for 30 min at 42°C followed by overnight incubation at 42°C with 10 µl of (denatured DIG-labelled) probe in 20 ml fresh EasyHyb solution. Following hybridization, the membrane was washed 2x (5min, 1x SSC, 0.1% SDS, 48°C) and 2x (15min, 0.5x SSC, 0.1% SDS, 65°C). The membrane was then allowed to cool to RT before 1x (wash buffer, 5 min) followed by incubation in blocking buffer for 30 min, RT (Roche). Anti-DIG AP-conjugate antibody (Roche) was diluted 1:10,000 in blocking buffer and incubated with the membrane for 30 min, RT. Non-specific antibody was washed 2x (wash buffer, 15 min, RT) and 1x (detection buffer, 5 min, RT). The membrane is the incubated in CSPD detection reagent (Roche) for 5 min, RT and finally for 10 min, 37°C. Membrane was exposed between 30 sec and 3 min using Hyperfilm (Kodak) in a Compact X4 developer.

qRT-PCR analysis of transcript depletion in conditional null mutants grown in non-permissive condtions and RNAi knockdown

Total RNAwas extracted from bloodstream form *T. brucei* using the RNeasy Mini Extraction Kit (Qiagen). Protocol was carried out according to manufacturer's instructions with a few deviations for *T. brucei*. Cells were centrifuged for 10 min, 800xgat RT in a 4K15 Sigma centrifuge, media was aspirated and the cell pellet was resuspended inbuffer RLT and β-mercaptoethanol was added at a 1:100 dilution.One volume of 70% ethanol was added to the lysate and the homogenous mixture was transferred to the provided column. RNA was bound to the column by centrifugation

for15 sec, 10, 000x*g*. The column was then washed with Buffer RWI and twice with Buffer RPE. Following the washes, the column was transferred to a sterile (RNAse free) Eppendorf tub, and the RNA was eluted in 50 µlRNase-free H_2O. The RNA concentration was then estimated by the A_{260}value of the nanodrop 2000c spectrophotometer (Thermo) with adjusted the path length setting (40) for RNA.cDNA was synthesized from the purified total RNA using the iScript cDNA Synthesis Kit (BioRad). A negative control reaction (without reverse transcriptase enzyme was performed in parallel to ensure the absence of gDNA in the samples. The components of the reaction were as follows.

Table 2.6: cDNA synthesis reaction components and amplification program

Reaction Mix (per 20 µl)
4 µl of 5x iScript reaction mix
template (500 ng total RNA)
1 µl iScript reverse transcriptase
up to 20 µl with nuclease-free H_2O

Amplification Program		
#of Cycles	T (ºC)	Time
1	25	5 min
1	42	30 min
1	85	5 min
	4	∞

Following RNA purification, and cDNA synthesis gene specific (Table. 2.8) mRNA levelsin were quantified by performing qRT-PCR using Sybrgreen master mix (Bio-Rad). The reaction was performed using a Bio-Rad iCycler Thermal Cycler.

Table 2.7: qRT-PCR reaction components and amplification program

Reaction Mix (per 20 µl)
10 µl of 2x SYBR Green reaction mix
3 µl of cDNA template (1:50 dilution)
0.6 µl forward primer (10 µM)
0.6 µl reverse primer (10µM)
up to 20 µl with nuclease-free H_2O

Amplification Program			
	#of Cycles	T (ºC)	Time
	1	50	2 min
	1	95	10 min
x40	1	95	15 sec
	1	60	30 sec
	1	72	30 sec
		4	∞

Table 2.8: List of primers for qRT-PCR of capping enzymes

Gene	Name	SequenceFunction	
TbCGM1	qRT-PCR – F	TCA ACG GGA AAG AAT CGT AGC A	qRT-amplicon-F
	qRT-PCR – R	CAC CAT ACC ATC GTT GTC GAG C	qRT-amplicon-R
TbCMT1	qRT-PCR – F	GGG AGC AAT TGC CAA AGC	qRT-amplicon-F
	qRT-PCR – R	CAC GCT TCG CCA GAA C	qRT-amplicon-R

T. brucei crude cell lysate for Western blot analysis of ectopic copy

Bloodstream form *T. brucei* cells grown to log phase density and $1x10^6$cells were harvested by centrifugation (800 x *g*, 10 min, 4°C) and washed twice in 1x TDB at 4°C. Next, cells were transferred to an Eppendorf tube and pelleted(10sec, 10,000x*g*, 4°C). The cells were then resuspended in 10 µl TDB buffer and lysed by addition of 4x sample buffer (Invitrogen). The cell lysate was heated 100°C for 10 min before loading onto a pre-cast Novex 4-12% Bis-Tris gels with MOPS running buffer (Invitrogen) at 200 V for 1.5 hours.

Table 2.9: 5x Trypanosome Dilution Buffer (TDB)

Chemical substance	Amount per 1L (g)
KCl	1.86
NaCl	23.4
$MgSO_4 \cdot 7H_2O$	1.23
Na_2HPO_4	14.2
$NaH_2PO_4 \cdot 2H_2O$	1.56
Glucose	18

Following protein separation by SDS-PAGE, the proteins were transferred onto a nitrocellulose membrane using theiBlot system (Invitrogen) on 'Program 3´ for 7 minutes. The membrane was then incubated for 5 min in Ponceau S (Sigma) to verify proteins transfer and then blocked in blocking buffer (50mM Tris Base, 0.15M NaCl, 0.25% (w/v) BSA, 0.05% Tween 20, 0.05% NaN_3, 2% fish skin gelatin (FSG), pH 7.4) for 1 h. The primary antibody (α-HA)(Millipore) was diluted 1:1000 in blocking buffer and incubated with the membrane in the SNAPid chamber (Merck) for 1 h at RT, rotating. After incubation, the membrane was washed 3x 35 ml wash buffer (1x PBS, 0.05% Tween-20) in the SNAPid vacuum manifold. IRDye(α-mouse)secondary antibody were diluted 1:15,000 in blocking buffer and incubated with the membrane for 1 h at RT, rotating. Membrane was imaged on the Licor system (700 channel – red (ladder), 800 channel – green (sample)).

Growth curve analysis of null and conditional null mutant *T. brucei*

For the growth curves the *T. brucei* BSF conditional null cells were washed 3x in TDB buffer to remove residual tetracycline from the media, and resuspended at either 5×10^4 cells/mL. Null cell lines were diluted without washing. Cells were counted every 2 days using the CasyCounter followed by dilution of each culture to 5×10^4 cells/ml (1 μg/ml tetracyclin was added every day to conditional null cells grown in permissive conditions). Growth curves were conducted in biological triplicates.

Mouse infectivity study

*Tb*CMT1 null and conditional null cells were subcultured and conditional null cells were grown with 1µg/ml tetracyclin. Groups of five (BALb/C) mice (dosed with and without doxycycline) were infected with 200 µl containing $2x10^5$ parasites by intraperitoneal injection (Fred Simeons and LasteStojanovski, University of Dundee). Doxycycline was added to the drinking water (0.2 mg/ml) in a 5% sucrose solution) of the group of animals that was to be dosed with for 1 week before infection and until the experiment was terminated. Infections were assessed by tail bleeding and cell counting in triplicate.

2.4Protocols for RNAi knockdown

RNAi knockdown construct generation, electroporation and phenotype analysis

A 535 bp fragment of *Tb*CGM1 was amplified by PCR and cloned into the $pRPa^{iSL}$ tagged locus, incudible stem-loop RNAi vector (a kind gift of the Horn Lab) with two sequential steps (Figure 2.2). The anti-sense strand was cloned using restriction endonucleases Acc651 and BamHI, and the sense strand was subsequently cloned using restriction endonucleases Bsp1201 and XbaI. Knockdown fragment was selected using (http://trypanofan.bioc.cam.ac.uk) on-line resource and the sequence for *Tb*CGM1 (Tb427.7.2080). This vector contains a tetracycline-inducible rRNA promoter which generates a stem loop RNAi transcripts. The resulting construct was linearized with AscI and $2.5x10^7$*T. brucei* 2T1 cells (another kind gift of the Horn lab) were electroportated with 10µg linearized construct which integrates into a specific rRNAspacer locus.Hygromycin (2.5 µg/ml, Sigma) was added 6 hours post-transfection and assayed for puromycin sensitivity. RNAi induced cells were grown in the presence of tetracycline (1µg/ml each day) and grown in parallel with uninduced cells. Every 24

hr cultures were counted and diluted to $1x10^5$. Gene knockdown was assessed by qRT-PCR using gene specific primers (Table 2.10).

Table 2.10:*Tb*CGM1 RNAi knockdown construct primers

Name	Sequence
TbCGM1-RNAi-F	GAT CGG GCC CGG TAC CGT CGA CTG GAG CCT AAC AGC
TbCGM1-RNAi-R	GAT CTC TAG AGG ATC CGT ACG CCC CAG TAA AGG TGA

2.5 Analytical methods for the analysis of nucleotide structure by mass spectrometry

TFA hydrolysis

Aliquots of 50 nmol of standards (GDP, GTP, m7GpppG) were hydrolysed by the addition of 250 µl freshly prepared 0.1 M TFA (20 min, 100°C). The reaction was stopped by flash freezing in liquid N2 and reaction products were freeze-dried in a freeze dryer overnight. Samples were then redissolved in 10 µl water for TLC analysis of Buffer A for analysis by LC-MS/MS (Table 2.11).

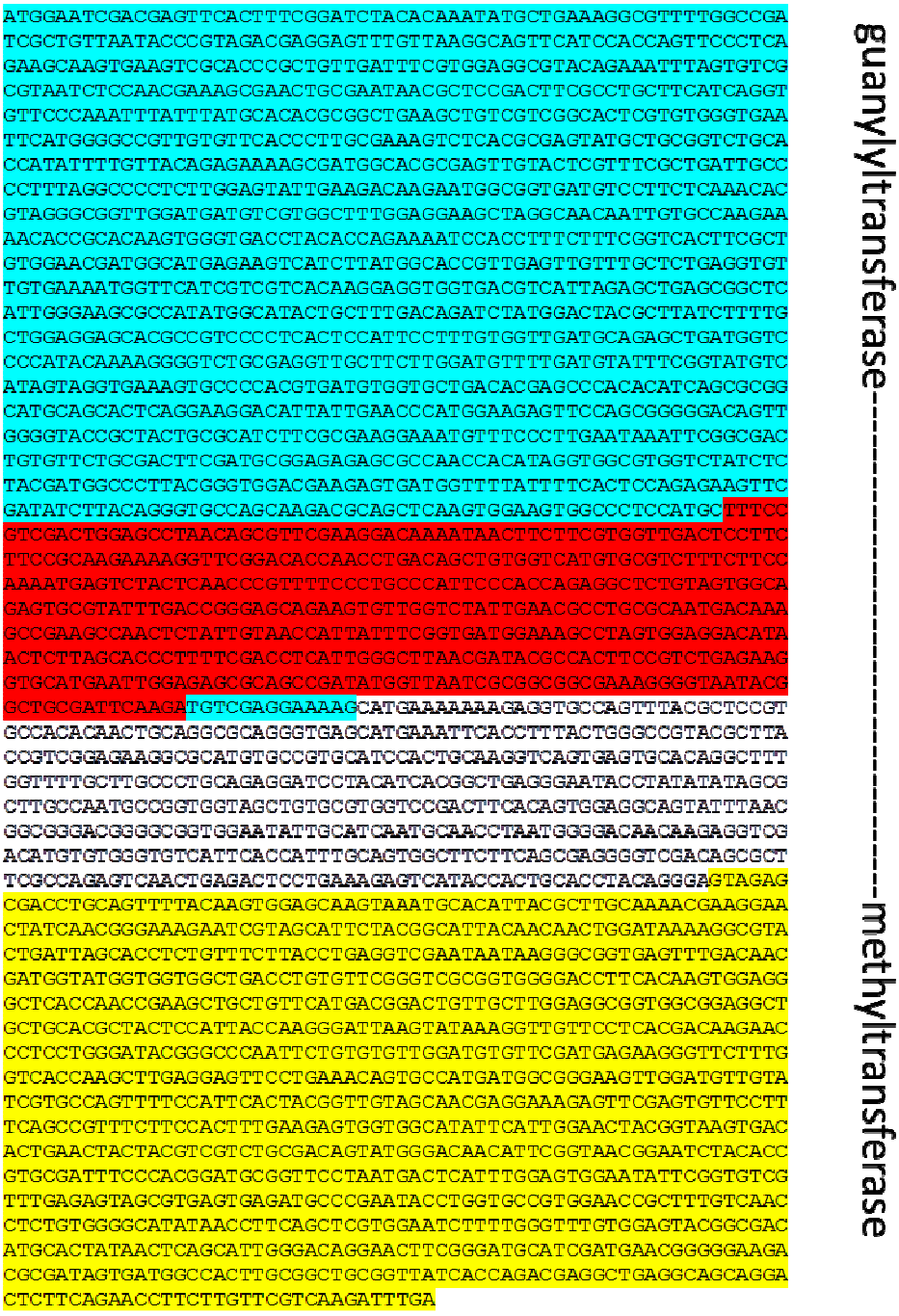


Figure 2.2: Sequence of TbCGM1. The bifunctional enzyme contains an N-terminal guanylyltransferase (blue) and a C-terminal methyltransferase (yellow). Gene fragment used for RNAi knockdown highlighted in red.

Table 2.11: HPLC buffer composition

	Chemical substance	Concentration
Buffer A	TBAA	10 mM
	Acetic Acid	15 mM
	Acetonitrile	2%
Buffer B	Acetonitrile	90%

TLC analysis

Thin layer chromatography was performed on 10 cm high PEI Cellulose-F (Merck) with 0.35M LiCl, 3.5M Urea as the mobile phase. Plates were by UV at 254 nm.

Electrospray mass spectrometry of cap structures

Initial experiments were performed on an Ultima Triple Quadropole mass spectrometer (Waters) in negative ion mode by direct infusion. Typical conditions were: Capillary voltage = -3kV, Cone voltage = -40V, Source temp = 110, Desolvation temp = 250, Desolvation gas = 500. The flow rate was 10 µl per minute, delivered using a 100 µl syringe (Hamilton) and a Harvard syringe pump. Spectra were collected, analysed and processed using Masslynx software (Micromass). For daughter ion scanning, conditions were as above with collision energy of 20V.

Later experiments were done on the QTRAP 4000 (Applied Biosystems) using Precursor ion (*m/z* 159) negative scanning mode with a nanospray ion source. Samples were also performed by direct infusion into an UltiMate 3000 HPLC (Dionex). Samples were separated on HPLC using a 40 min gradient from 0-90% buffer B. Typical MS/MS conditions were: mass range = 360-600, over 10 sec. Declustering Potential (DP) = -70, collision cell entrance potential (EP) = -8, collision cell exit potential = -12, collision energy = -30V. Curtain gas = 30, Collision gas = 11, IonSpray voltage = -2000, Ion Source gas I = 25, interface heater temp = 175 °C.

HEK-293 cell harvest and total RNA purification

HEK-293 WT cells were a kind gift of the MRC tissue culture facility, University of Dundee. Media was aspirated from a flask of confluent HEK-293 WT cells, and the cells were washed once in 5 ml of PBS, centrifuged (1700 RPM, 5 min, 4ºC) and the pellet was resuspended in 1 ml TRIzol Reagent (Ambion). Homogenized cells could be stored at -80ºC for up to 1 month. Once thawed from frozen, 200 µl of chloroform was added to the lysate and votexed for 15 sec. The mixture was then centrifuged to separate the phases (13,000 x g, 15 min, 4ºC) and 475 µl of the aqueous phase was transferred into a new RNAse-free tube. It is important not to contaminate the RNA-containing aqueous phase with the interphase portion of the homogenized sample when pipetting. The RNA is precipitated by the addition of 500 µl of 100% isopropanol to the aqueous phase isolated in the previous step, incubated at RT for 10 min and centrifuged (12,000 x g, 10min, 4ºC). To wash the RNA pellet, the supernatant was removed, replaced with 1 ml of 75% ethanol, centrifuged (7500 x g, 5 min, 4ºC), and the wash discarded. The RNA pellet was then air dried for 10 min, resuspended in 30 µl diethylpyrocarbonate (DEPC) water, and stored at -80ºC.

2.6 Polysome profile analysis

Polysome fractionation and RNA extraction

Procyclic form *T. brucei* were grown to log phase growth at 28°C without CO_2 in SDM-79 medium, and bloodstream form *T. brucei* were grown to log phase growth at 37°C with 5% CO_2 in HMI9-T medium. Polysomes were prepared by resolving a native cytosolic extract through a sucrose gradient. Cultures of *T. brucei* cells were incubated with 50 µg/ml cycloheximide for 10 min prior to the start of polysome purification procedure. Cells were pelleted by centrifugation at 4°C and PCF cells were washed with

PBS containing cycloheximide (137 mMNaCl, 2.7 mMKCl, 10 mM Na_2HPO_4 • 2 H_2O, 2 mM KH_2PO_4, 1 mg/ml cycloheximide, pH 7.4), while BSF cells were washed with TDB containing 1mg/ml cycloheximide. Cells were resuspended in polysome lysis buffer (120 mMKCl, 2 mM $MgCl_2$, 20 mMTris pH 7.5, 1 mM DTT, 1% n-octylglycoside, 50 µl RNAsin, 2 µg/ml leupeptin, 1 µg/ml aprotinin, 1 µM TLCK, 1mM PMSF, 1mg/ml cycloheximide). The detergent n-octylglycoside(NOG) was chosen because it does not absorb at 254 nm. The lysates were loaded on top of 10 ml sucrose gradients (5 increments, 2ml each: 10%–50%) and centrifuged for 2 h at 38,000 rpm at 4°C in a Beckman ultracentrifuge using a SW41Ti rotor. Gradients were fractionated (0.5 ml fractions) and analysed for nucleic acid content by a Nanodrop spectrophotometer at 254 nm. RNA was purified using an RNeasy kit (Quiagen) from each fraction and from pooled subpolysome and polyribosomal fractions. Gradient analysis was also performed using a gradient collector (Teledyne) with continuous monitoring at 254 nm. Individual fractions were collected with a Foxy Jr. (Teledyne) fraction collector want RNA was purified as above. The RNA was quantitated and checked for purity by Nanodrop spectrophotometer and used for qRT-PCR analysis or RNA-seq cDNA library preparation.

Polysome fractionation and RNA extraction following chemically induced ER stress

Procyclic form *T. brucei* were grown to log phase growth at 28°C without CO_2 in SDM-79 medium, and bloodstream form *T. brucei* were grown to log phase growth at 37°C with 5% CO_2 in HMI9-T medium. Cells were either treated with 0.8 µg/ml tunicamycin or DMSO as a solvent control for 2 hr at 37°C for BSF or 28°C for PCF *T. brucei*. Polysomes were prepared as described above, with the exception that 2% NOG was used in the lysis buffer. In a separate experiment, cells were treated with 4 mM DTT for

1 hr at 37°C for BSF or 28°C for PCF *T. brucei*, and polysomes were prepared as described above.

Table 2.12: qRT-PCR primers

Gene Name	qRT-PCR Forward Primer	qRT-PCR Reverse Primer
TbTransketolase	AAG GCT TCG GTG CCT CTG CG	AAT GGG ACC GCT TCG CTG GC
TbAconitase	CGC GGC TGC TGT ACT ATC GGG	ACT CGC CCA GCC AGG GCA TA
TbmMDH	AGC TGC CCG GTG GAA CTT GG	CGA AGG CAA AAC CTG CTG CGA
TbCap5.5	CTG CAG CTG AAG AAG TGC CCG T	CCT GCG GCT GCG GTA CGT C
TbRab18	GGT TCC GGA CGC TCA CCA GC	CGC CTC CTC CAC CCC ATC CA
TbDPMS	CTG CGC GCA CTG TAG GCG AGA	GAG TAG AAC CGT GAG CGC GGT G
TbBiP	CTA CCC ATG CGG ACA ACC	CAG CAG GAG GAA TTC CCG AC
TbGlucosidase	GTC ATT CTC GGG TAC GAA GG	GGT TCC GCA GTA CAA GGG AC
TbCalreticulin	GAT GGA CCC AGC TGA ACT C	GTA ACT TGG GCG CTT CTT CC
TbPDI	GGT CAA GCC TAA TGA AAC TGC	CAC CGA AGT CCC TCC AAA CC
TbUGGT	CTG GAG GTA AGG CTG ACG	GAG TCG CAT GGT TTG C
TbGAPDH	TCG TGA TGG GCG TGA ACC AC	GTG GAG ATG CCG AAG CCC TC
TbEnolase	AGG ATG CTG TGA ACG TTG GT	TCC ATG CAG ATG GCG AAC TT
TbTrypanothione Reductase	TGC AGA CTT CGC ATG GAC CG	GCA GAC TCG CGA AGG TGG TC
TbPRP	GAA GTT GAA GGT GTT GTG AGT CC	CCT CCA GCG TGA TAT CTG TTA CC
TbTERT	GAG CGT GTG ACT TCC GAA GG	AGG AAC TGT CAC GGA GTT TGC

Preparation of cDNA libraries for RNA-seq

Total RNA, subpolysomal, and polyribosomal RNA was isolated from BSF and PCF *T. brucei* followed by poly(A) mRNA enrichment. Three methods of mRNA enrichment were tested: streptavidin poly-T oligo-attached magnetic beads (Dynabeads, Invitrogen), poly-T oligo-attached magnetic beads (Illumina), and Terminator 5′-Phosphate-Dependent Exonuclease (Epicentre). The mRNA was then fragmented into 200 nt fragments using Covaris Adaptive Focused Acoustics process. Operating conditions - sample volume: 130 µl, duty cycle: 10%, intensity: 5, cycles per burst: 200, processing time: 60 s, water bath temperature: 4°C, power mode: frequency sweeping, degassing mode: continuous. Fragmented mRNA was concentrated by ethanol precipitation and measured on a RNA Pico chip (Agilent 2100 Bioanalyzer). All enzymes and buffer

purchased from NEB. The first strand of cDNA was synthesized using reverse transcriptase and random primers, followed by second strand cDNA synthesis which removes the RNA template producing double-stranded cDNA. To blunt-end the DNA fragments, an end repair reaction was performed with Klenow polymerase, T4 DNA polymerase, and T4 polynucleotide kinase. A single 3´ adenosine overhang was added to the cDNA allowing the ligation of Illumina adaptors. These adaptors contain primer sites both for sequencing and complimentary annealing onto the Illumina flow cell surface. Adaptor ligated cDNA fragments were measured on an Agilent DNA chip. The final cDNA library was sequenced on a MiSeq Personal Sequencer of HiSeq2000 (Illumina). Following sequencing, alignment and quantitation were performed using the following workflow:

Bioinformatic analysis of RNA-seq data (Bernardo Foth/Michele Tinti)

MiSeq RNA-sequencing: Sequence reads were mapped to the 927 reference genome using Tophat and filtered for minimum mapping quality (-q 30) and to softclip the shorter half of mappings when a read is mapped across an intron. Sequence reads per exon were counted using BEDtools(annotation provided as GFF file). Next, the total read count per gene was calculated by summing up the reads over all exons belonging to the same gene. Differential expression analysis was done using the R module edgeR. adjusted p value (adj.P.Val) indicates how significantly differentially "expressed" a particular gene in a particular pairwise comparison is. The adjusted read counts "pseudocounts" per gene following normalisation, give a feel for the reproducibility between the biological replicates and for absolute transcript levels per gene.

High-Seq RNA sequencing analysis was conducted using the same pipeline as above with the following amendments. Differential analysis done with edgeR with adjusted p-value < 1e-10. A different statistical model (GLM = generalized linear model) was used

within the previously used software package edgeR, which considers "all data at once" (as opposed to only considering the data for a given pairwise comparison at a time).

RESULTS – CHAPTER I

RESULTS – CHAPTER I

3. Essentiality of the mRNA capping enzymes

3.1 Generation of *Tb*CMT1 conditional null mutants and investigation of their phenotype

For all cells in which it has been investigated, the presence of the 5´RNA cap is vital for cell viability, effective gene expression, and mRNA stability. However, the question of cap essentiality remains unanswered in bloodstream form *T. brucei.* Cap formation occurs in three steps: first, the 5´ triphosphate group on RNA is hydrolysed by a triphosphatase to produce diphosphate–RNA. Next, the guanylyltransferase catalyses the addition of GMP to the diphosphate–RNA to produce the guanosine cap. Finally, the cap is acted upon by a guanine-7-methyltransferase, catalysing the methylation of the guanosine cap at the N-7 position. In addition to the three capping enzymes described, trypanosomatids have an additional guanylyltransferase and methyltransferase in the form of a bifunctional enzyme (Takagi et al., 2007).The significance of the bifunctional capping enzyme remains unclear, so we propose to compare the essentiality of this bifunctional enzyme (*Tb*CGM1) with the monomeric methyltransferase (*Tb*CMT1) in BSF *T. brucei*.

All conditional null mutants were generated in the *T. brucei* Lister 427 bloodstream form cell line which has been genetically modified to express the tet repressor protein (TetR) and T7 RNA polymerase under the selection of neomycin and is therefore known as single marker (SM) (Wirtz et al., 1999). Using this cell line, conditional null mutants were created by sequential allelic replacement with antibiotic resistance cassettes by homologous recombination. In case of the gene of interest is essential for the parasite's viability, an ectopic copy of the gene was introduced after replacement of

the first allele. The ectopic copy is incorporated into the rDNA locus by homologous recombination using a tetracycline inducible promoter to drive expression of the gene.

A schematic of the strategy used to construct *TbCMT1* gene replacement plasmids is shown in (Figure 3.1). Upstream and downstream *TbCMT1* UTRs were first amplified by PCR and sequenced to ensure sequence differences between the TritrypDB annotated *T. brucei* 927 strain and the laboratory *T. brucei* 427 strain did not interfere with primer design. About 500 bp of 5′ and 3′ UTRs containing the necessary restriction enzyme sites were amplified and then joined together by PCR via a pre-determined complimentary sequence. The 5′ forward and 3′ reverse primers were then used to amplify the single piece of DNA containing both UTRs and cloned into the pGEM5Zf vector. An antibiotic resistance cassette was then introduced in place of the complementary sequence, and the resulting plasmid was sequenced to confirm correct amplification and insertion. The final knockout construct was released from the vector with *NotI* digestion and it was then ready for transfection.

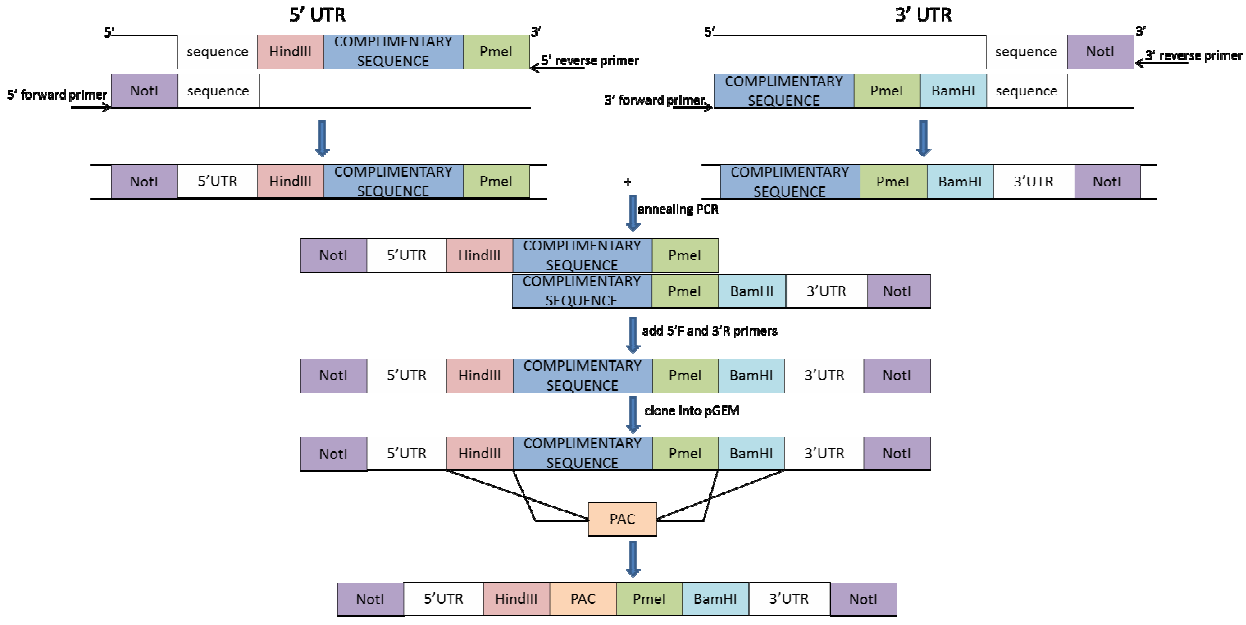


Figure 3.1: Strategy for the generation of gene replacement constructs.

Homologous recombination, targeted to the 5′ and 3′UTRs described above, allowed the first and second alleles of the *TbCMT1* gene to be replaced with PAC and HYG resistance cassettes, respectively. Given the possibility that the gene might be essential *in vitro*, before the second allele was replaced, a tetracycline-inducible ectopic copy of the *TbCMT1* gene in the plew100 vector was introduced and induced to maintain cellular protein levels during the second round of allelic gene replacement (Figure 3.2). The genotypes of the recovered cells were analyzed by Southern blot, showing the loss of the endogenous *TbCMT1* alleles and appearance of the ectopic *TbCMT1* copy in the conditional null mutant (Figure 3.3, panel A). These Southern blots also confirmed the antibiotic resistance cassettes were in the correct loci (Figure 3.3, panels B and C). When Wirtz and colleagues generated the SM cell line described above, a fragment of the hygromycin gene used to target the TetR construct remained in the genome. Consequently, an additional band was present at about 1.3 kb band on the Southern blot probed with HYG (Figure 3.3, panel C).

To assess the phenotype of the *Tb*CMT1 conditional null mutant, cell growth rate was observed under permissive (+tet) and non-permissive (-tet) conditions over 10 days (Figure 3.4). After 24h in non-permissive conditions, less than 1% of *Tb*CMT1 transcript was detected compared with cells grown permissive conditions (Figure 3.4, panel A). The growth curve for the mutants showed that non-permissive conditions did not substantially impair the growth of the cells, indicating *Tb*CMT1 is dispensable for *T. brucei* survival *in vitro* (Figure 3.4, panel B).

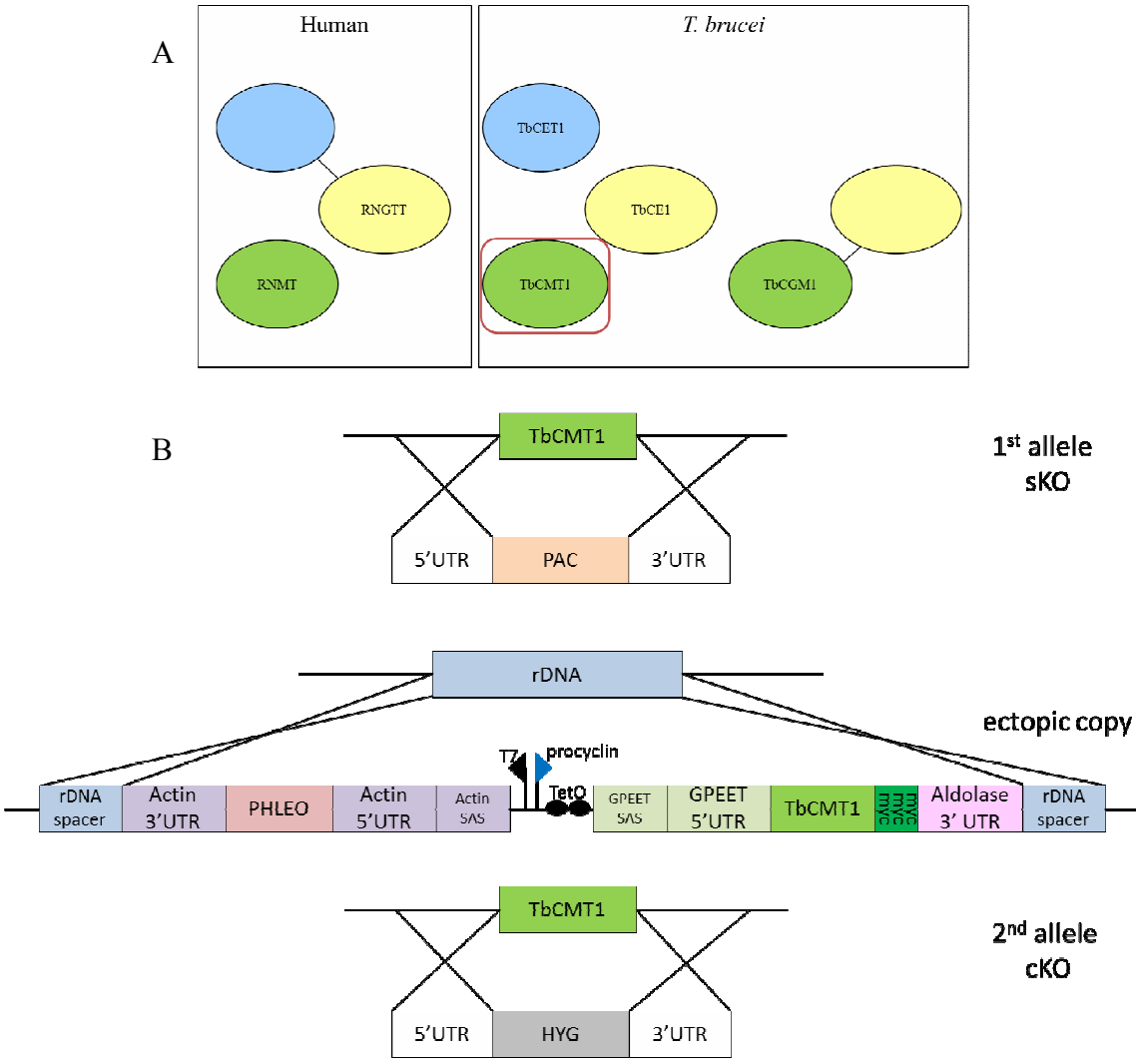


Figure 3.2: Bloodstream form *T. brucei TbCMT1* gene replacement strategy. (A) Cartoon representing the arrangement of human and *T. brucei* mRNA capping enzymes, red box outlines monomeric methyltransferase enzyme *Tb*CMT1. (B) The first allele of *TbCMT1* was replaced with a PAC resistance cassette. An ectopic copy of a tetracycline-inducible version of *Tb*CMT1 with three C-terminal c-myc tags was introduced into the rDNA locus using a procyclin promoter under the control of two tet-operator (TetO) sequences. Selection was via a phleomycin (PHLEO) resistance marker driven by a T7 promoter. Activation of the ectopic copy with tetracycline allows replacement of the second allele, regardless of essentiality, with a HYG resistance cassette.

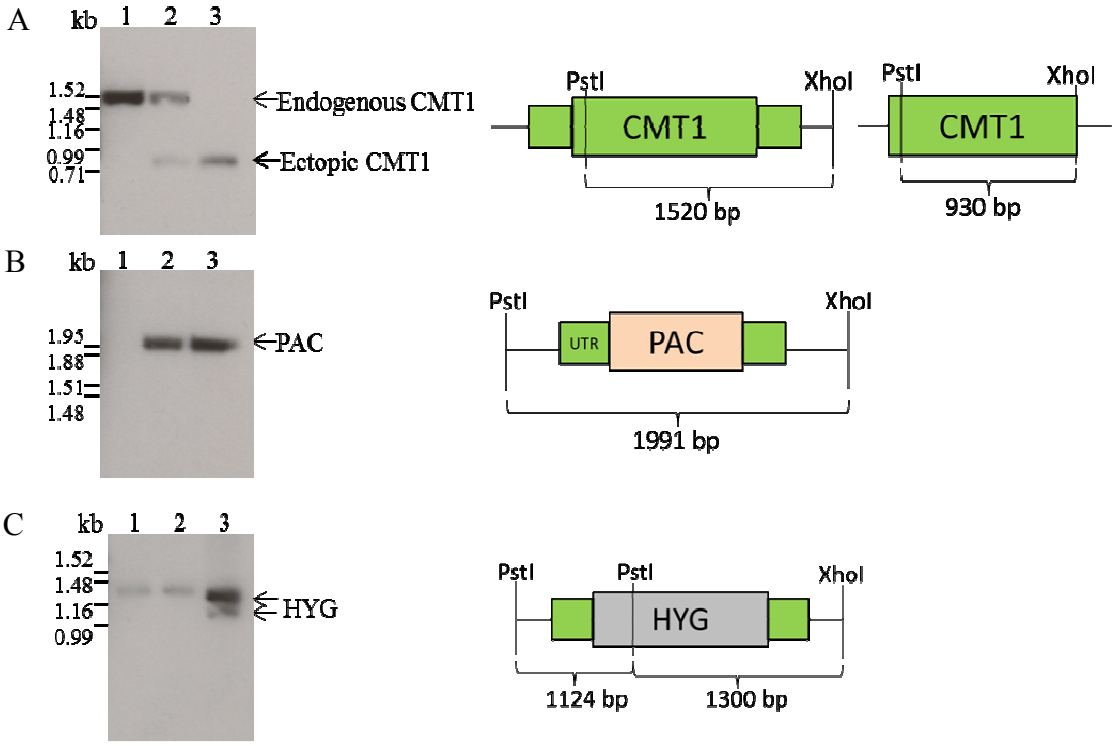


Figure 3.3: Southern blot analysis of the bloodstream form *Tb*CMT1 conditional null mutant. Genomic DNA was extracted from parental SM (*Tb*CMT1$^{+/+}$) cells (lane 1), cells with one *Tb*CMT1 allele replaced with PAC and an ectopic tetracycline inducible copy of *Tb*CMT1 introduced (*Tb*CMT1$^{-/+}$, *Tb*CMT1^{Ti}) (lane 2), and the same cells following replacement of the second *Tb*CMT1 allele with HYG (*Tb*CMT1$^{-/-}$, *Tb*CMT1^{Ti}) (lane 3). The DNA was digested with restriction enzymes PstI and XhoI. The maps show the predicted gDNA fragment sizes for CMT1 and for the correct replacement of the antibiotic resistance genes. The Southern blots were probed with A: CMT1 ORF, B: PAC, C: HYG. Note the weakly hybridizing band to the HYG probe in lanes 1 and 2 is due to the fragment of the hygromycin gene in the SM cell line.

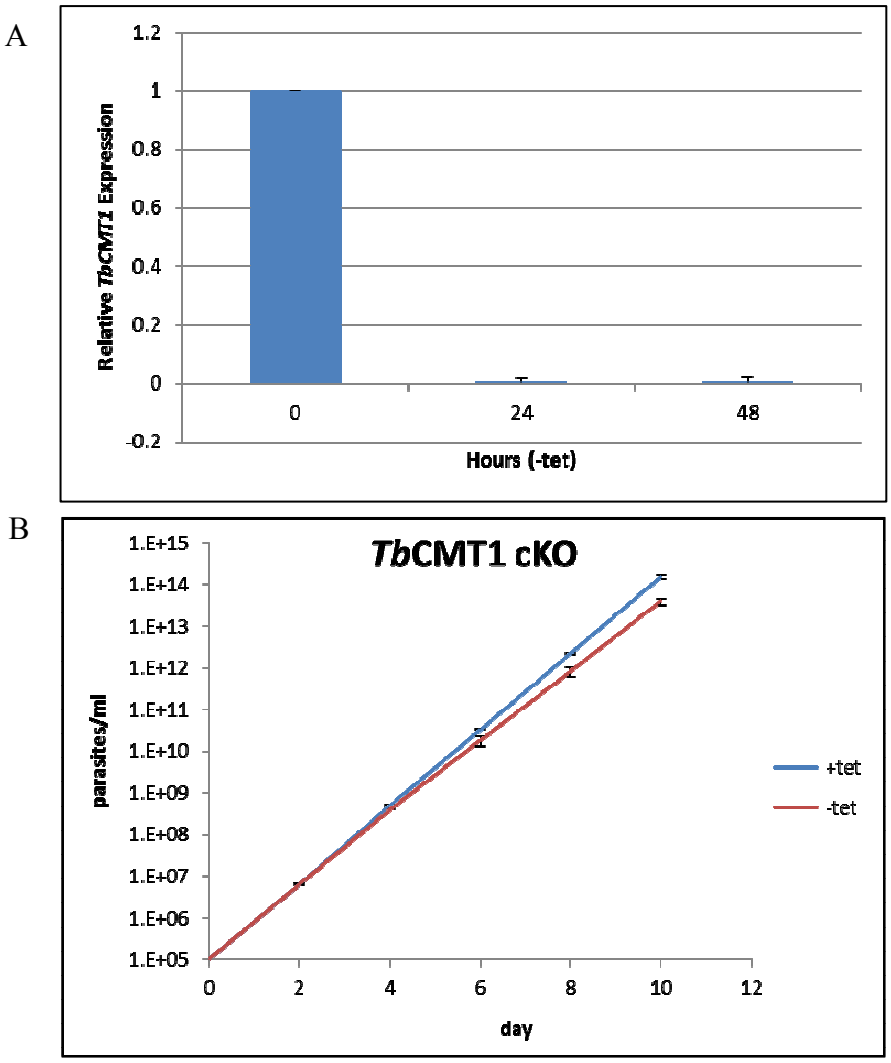


Figure 3.4: *Tb*CMT1 is not essential for bloodstream form cell growth *in vitro*. (A) Total RNA cells grown in non-permissive conditions for 48 hours was purified and analysed by qRT-PCR to analyse the downregulation of TbCMT1 transcript. Three biological replicates were analysed; the mean ± standard deviations are shown. Results were normalized to the constitutive gene: telomerase reverse transcriptase (TERT) (Brenndörfer and Boshart, 2010). (B) Growth curve of TbCMT1 cKO cells grown in permissive (+ tet) and non-permissive (- tet) conditions. *Tb*CMT1 conditional null mutants (+/- tet) were grown in parallel; cultures were counted every 24 h and diluted to $1x10^5$ cells/ml every two days. Three biological replicates were counted in triplicate; the mean ± standard deviations are shown.

3.2 Generation of *Tb*CMT1 null mutants

As further and final validation that *Tb*CMT1 is not essential for parasite survival in culture, a *TbCMT1* null mutant (without an ectopic copy) was created. The *TbCMT1* single allele replacement (*Tb*CMT1$^{-/+}$) cells described above were electroporated and the second allele replaced with a HYG resistance cassette. Recovered cells were analysed by Southern blot and the disappearance of the 528 bp band corresponds to the *TbCMT1* ORF in lane 3 (Figure 3.5, panel A) and the presence of the *PAC* and *HYG* genes in the correct loci, indicated by the presence of the 2 kb and 1.1 and 1.3 kb bands in lane 3, respectively, confirmed the generation of the *TbCMT1* null mutant (Figure 3.5, panel B and C). As described in section 3.1, an additional band is seen in the blot probed with *HYG* (at 1.3 kb) due to the residual fragment of the gene in the gDNA of the *T. brucei* SM cell line (Figure 3.5, panel C).

3.3 Investigation of *Tb*CMT1 null mutant phenotype

To assess the viability of *Tb*CMT1 null mutants, the cell growth rate was observed in comparison with single marker cells over 10 days (Figure 3.6). The growth curve for the mutants showed that the absence of *Tb*CMT1 did not substantially impair the growth rate of the cells, confirming *Tb*CMT1 is not essential for *T. brucei* survival *in vitro*. From these results, we decided to investigate whether *Tb*CMT1 influences *T. brucei* parasitemia in animal infections. Three groups of mice were compared in the study: single marker cells as a positive control for the infection, *Tb*CMT1 null mutant cells, and *Tb*CMT1 conditional null mutant cells. For the week prior and throughout the course of the infection, doxycycline was added to the drinking water of the mice infected with the *Tb*CMT1 conditional null mutant to induce expression of the ectopic copy.

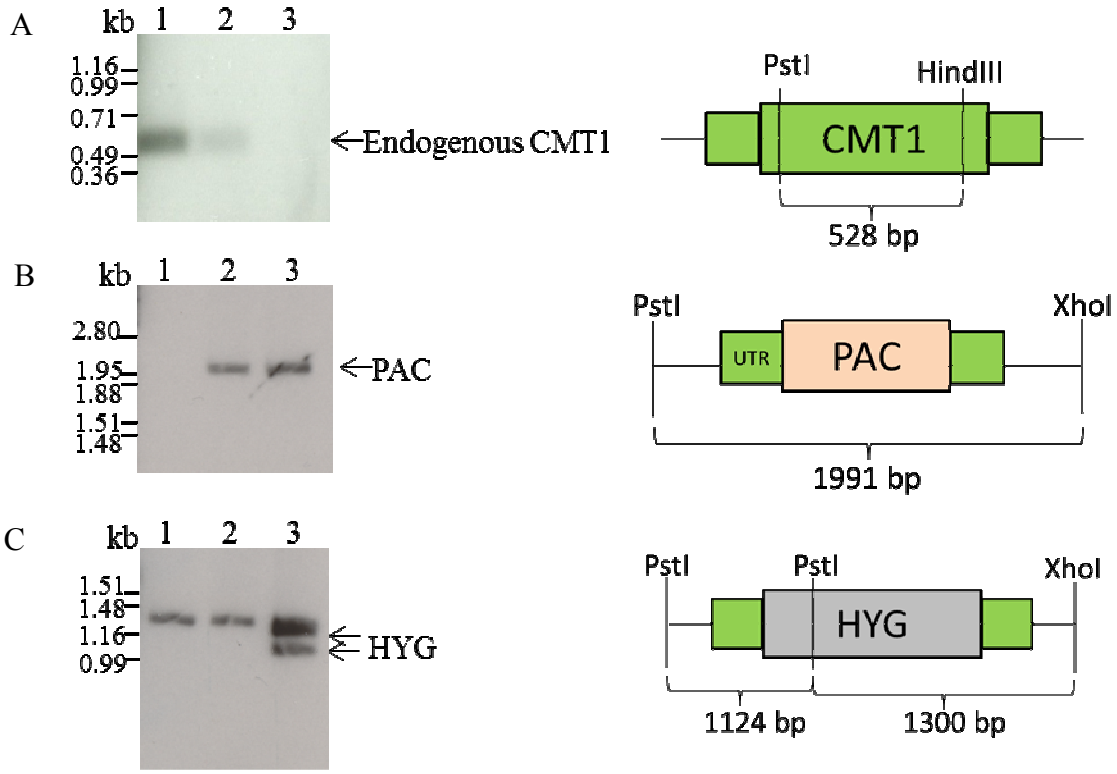


Figure 3.5: Southern blot analysis of the bloodstream form *Tb*CMT1 null mutant. Genomic DNA was extracted from parental SM (*Tb*CMT1$^{+/+}$) cells (lane 1), cells with one *Tb*CMT1 allele replaced with PAC (*Tb*CMT1$^{-/+}$) (lane 2), and the same cells following replacement of the second *Tb*CMT1 allele with HYG (*Tb*CMT1$^{-/-}$) (lane 3). The DNA was digested with restriction enzymes PstI and XhoI. The maps show the predicted gDNA fragment sizes for CMT1 and for the correct replacement of the antibiotic resistance genes. The Southern blots were probed with A: CMT1 ORF, B: PAC, C: HYG. Note the weakly hybridizing band to the HYG probe in lanes 1 and 2 is due to the fragment of the hygromycin gene in the SM cell line.

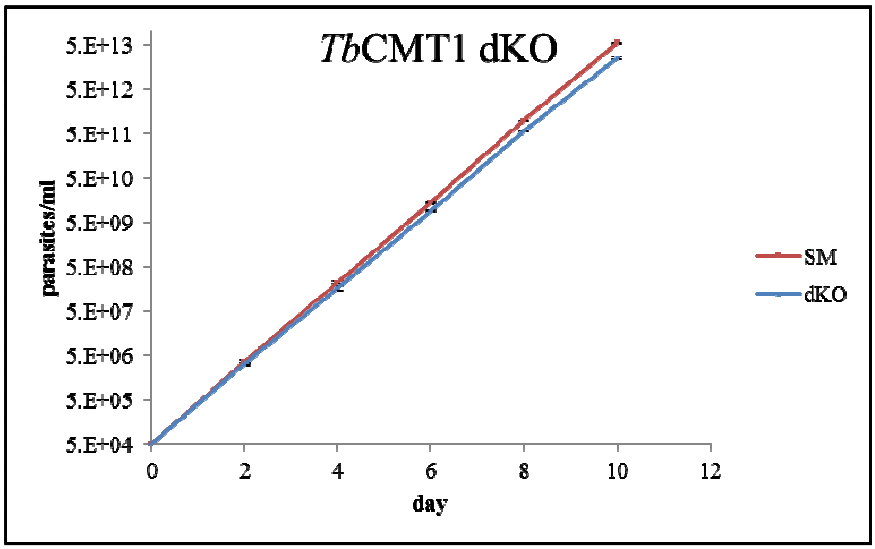


Figure 3.6: *Tb*CMT1 is not essential for bloodstream form cell growth *in vitro*. Growth curve of TbCMT1 null mutant (dKO) cells. SM and *Tb*CMT1 null mutants were grown in parallel;cultures were counted every 24 h and diluted to $5x10^4$ cells/ml every two days. Three biological replicates were counted in triplicate; the mean ± standard deviations are shown.

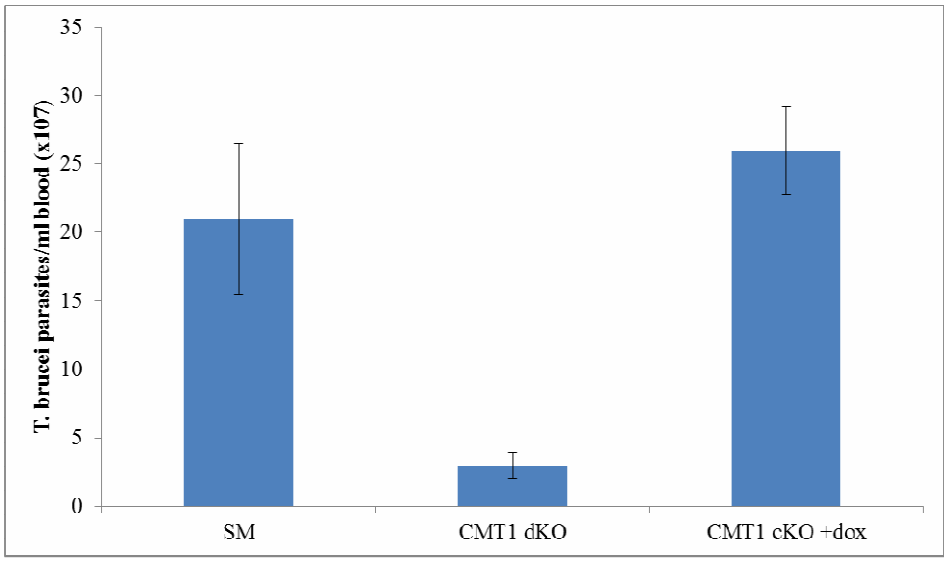


Figure 3.7: *Tb*CMT1 is important for *T. brucei* virulence *in vivo*. Mice were infected with SM, *Tb*CMT1 null (dKO), and *Tb*CMT1 conditional null (cKO) cell lines and blood parasitemias were measured after 3 days of infection. TbCMT1 cKO mice were dosed with doxycycline for 7 days before and throughout the infection. Ten animals (nine for SM) were counted in triplicate, the mean ± standard error are shown.

The SM cell lines grew well, as expected, but the parasitemia in *Tb*CMT1 null infected mice was very low. In the presence of the doxycycline induced ectopic copy, the infection was as high as SM cells suggesting the decrease in parasitemia was a direct result of the absence of *Tb*CMT1 (Figure 3.7).

Trypanosomes predominantly regulate gene expression post-transcriptionally; therefore, we decided to investigate the effect of the absence of *Tb*CMT1 on translation using polysome profiling. Cell lysates from cultured single marker and *Tb*CMT1 null mutant cells were fractionated on sucrose gradients (10-50%) in order to compare the actively translated mRNAs between cell lines. The polysomal material was equally abundant in single marker and *Tb*CMT1 null mutant cells. Although it is possible that the distribution of certain transcripts within the polysomal pool is different between the cell lines, an overall translation defect was not seen in *Tb*CMT1 null mutant cells (Figure 3.8).

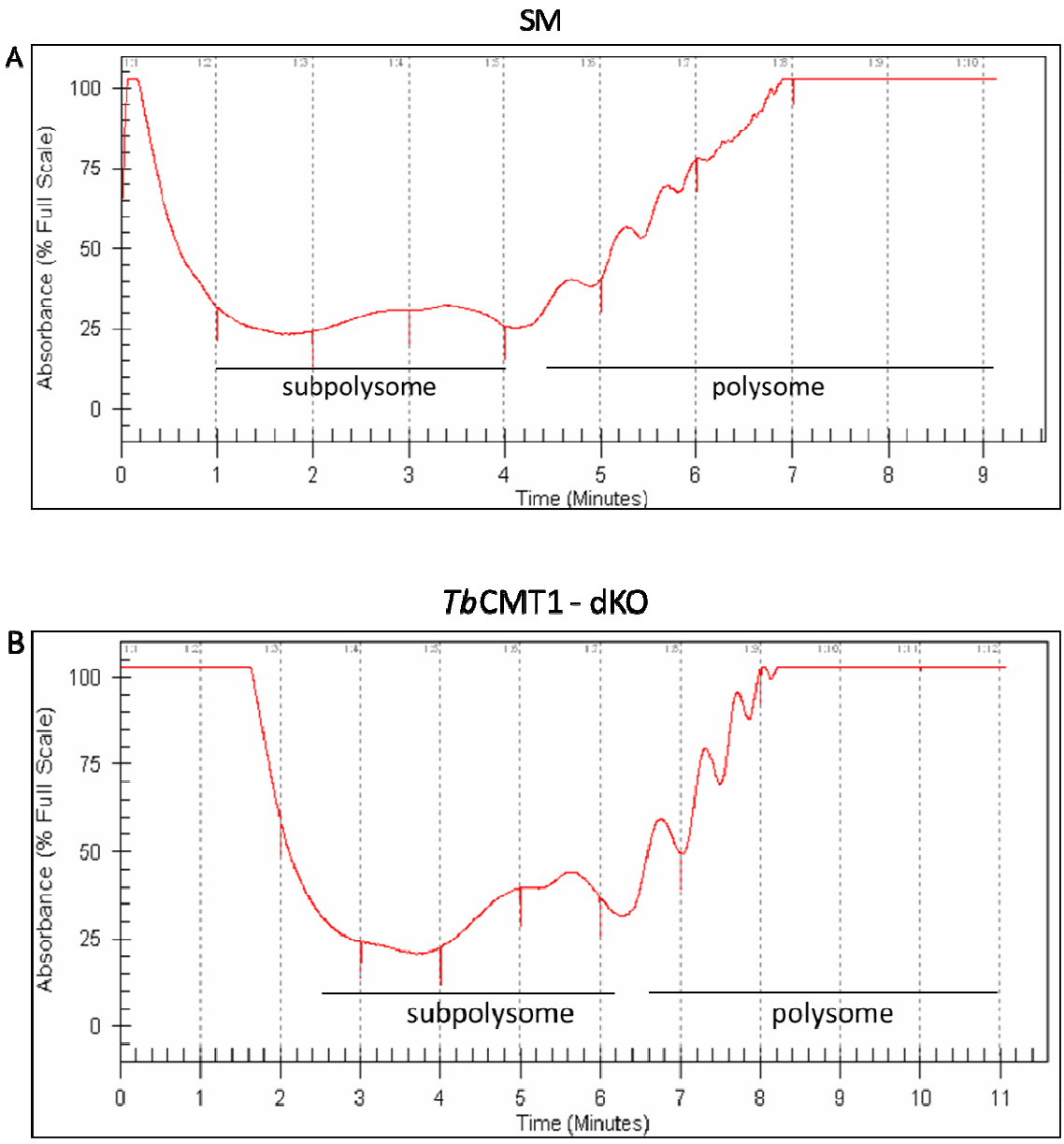


Figure 3.8: Analysis of the ribosome distribution in single marker and *Tb*CMT1 null *T. brucei*. Polysome profiles of (A) single marker (SM) and (B) *Tb*CMT1 null mutant (dKO) cells. Absorbance at 254 nm, an indicator of RNA abundance, (y axis) is plotted against the sucrose gradient fractions (x axis). The positions of free ribosomal subunits and polysomes are indicated. Polysomal fractions 8-11 are unresolved due to the RNA abundance in the samples reaching the limit the absorbance detection range of the gradient fraction collector.

3.4 RNA-seq analysis of SM and *Tb*CMT1 null mutants

To gain more insight into the mechanisms underlying the decreased virulence of *Tb*CMT1 null mutants, we performed RNA-seq on SM and *Tb*CMT1 null mutant (dKO) cell lines. Four biological and two technical replicates were sequenced using an IlluminaMiSeq sequencer. A high correlation for the read counts of biological replicates for SM (average Pearson correlation coefficient R = 0.938) and *Tb*CMT1 dKO (average R=0.97) *T. brucei* confirmed the reproducibility of the methodology (Figure 3.9). The differential expression of genes affected by the loss of *Tb*CMT1 was evaluated using three different statistical approaches and the common hits of all three were used for further analysis (Figure 3.10). The MA plot visualizes the relationship between intensity and the differences between the two data sets. We detected 29 significantly differentially expressed genes which exhibited a 2-fold or greater change ($P<0.05$) (Table 3.1). Of these genes, 21 genes (72%) were downregulated in *Tb*CMT1 null cells compared to SM control and 8 genes (28%) were upregulated. Of the downregulated genes, the majority, 11 genes (52%) belong to the cysteine peptidase family of proteins. Other observed categories of differentially expressed genes include VSG and expression site associated genes (ESAGs), and novel coding sequences that are as yet unannotated in the genome. Although at first glance it appeared that *Tb*CMT1 was the only third most downregulated transcript in the dataset, upon closer analysis, the RNA-seq coverage plot revealed the sequence reads assigned to gene were from the 5′ and 3′ UTRs. None of the reads were detected from the ORF, confirming the generation of a *Tb*CMT1 null cell line (Figure 3.11).

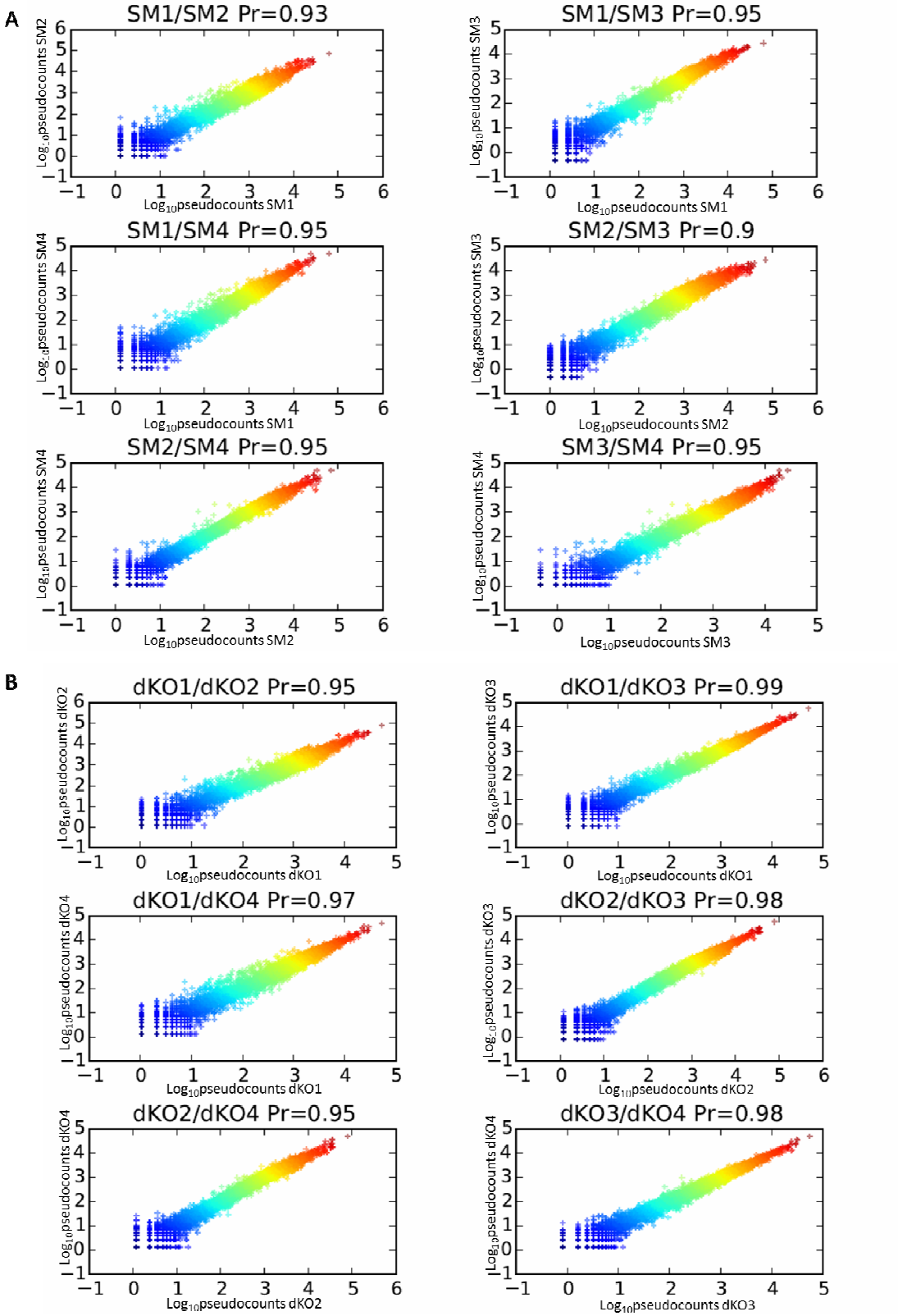


Figure 3.9: RNA-seqbiological replicate analysis. Matrices of pairwise scatter plots of the correlation between log_{10}normalized read counts between two biological replicates of (A) single marker (SM) and (B) TbCMT1 null (dKO) *T. brucei* samples. Pairwise Pearson correlation values (Pr) are displayed above each plot. Graph generated by Michele Tinti, MAJF lab.

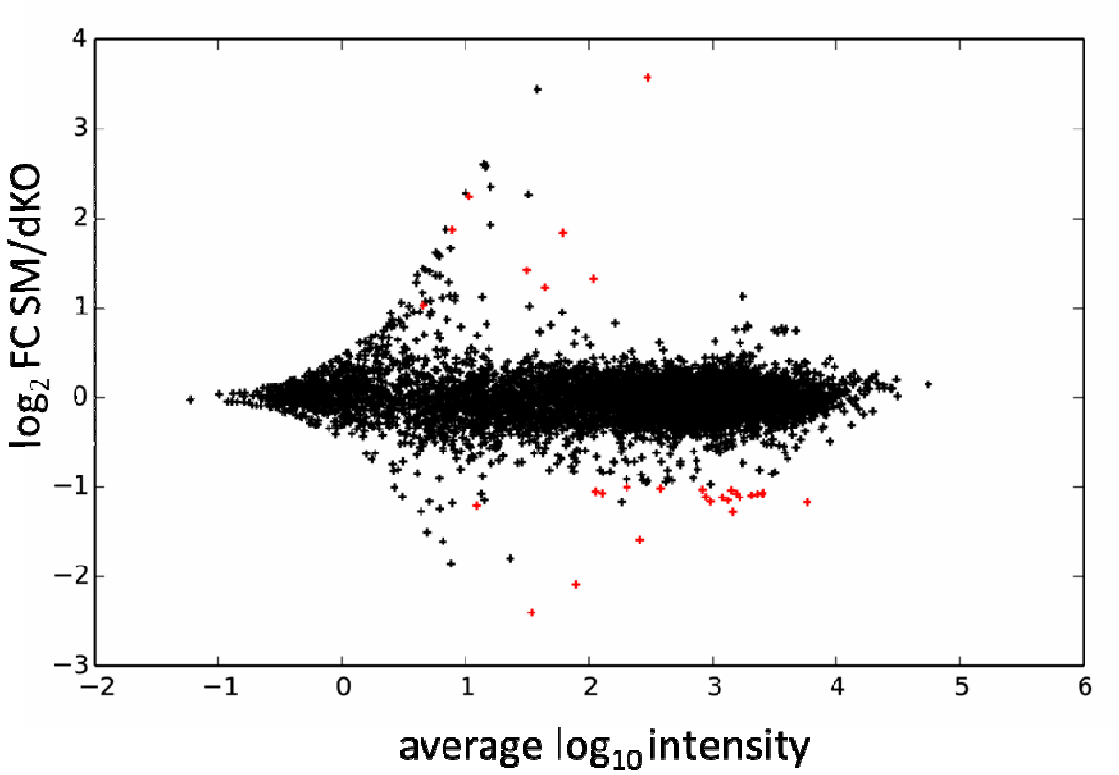


Figure 3.10: MA plot of RNA-seq dataset. An MA plot creates a visual representation of differentially expressed genes identified in SM and *Tb*CMT1 null (dKO)*T. brucei* cell lines. Data represent individual gene responses derived from DEseq differential expression analysis plotted as average intensity vs the $\log_2$ fold-change (p-values <0.05). A negative fold change represents down-regulated genes and a positive fold change represents up-regulated genes. Red points designate genes that were found to be significantly differentially expressed in all three processing algorithms. Graph generated by Michele Tinti, MAJF lab.

Table 3.1: Summary of RNA-seq results. Genes with a minimum 2 fold change in expression between SM and *Tb*CMT1 null (dKO) *T. brucei* cell lines(p-values <0.05). VSG and expression site associated genes (ESAG) are highlighted in blue, cysteine peptidases are highlighted in green, and genes with unannotated coding sequences (CDS) are highlighted in purple. The new CDS failed to reveal any conserved

	Gene ID	log2FC	Gene Description
down-regulated in dKO	Tb927.9.257	-2.403682165	variant surface glycoprotein (VSG, pseudogene), putative
	Tb927.11.15180	-2.085335054	hypothetical protein
	Tb927.10.4500	-1.588245791	**TbCMT1 - mRNA capping methyltransferase, putative**
	Tb927.6.970	-1.27540401	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.11.20590	-1.209494073	variant surface glycoprotein (VSG, pseudogene), putative
	Tb927.6.1020	-1.170643006	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (Rhodesain)
	Tb927.6.960	-1.162713813	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.6.980	-1.148629865	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.6.1060	-1.115541708	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.6.990	-1.115291633	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.6.1010	-1.10928205	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.6.1000	-1.093595841	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.11.1110	-1.081031212	**calpain, putative**
	Tb927.6.1040	-1.079753799	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	Tb927.6.1050	-1.075734817	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	XLOC_004762	-1.068141941	cufflinks prediction (new CDS)
	Tb927.6.1030	-1.06450382	cysteine peptidase precursor,cysteine peptidase, Clan CA, family C1, Cathepsin L-like (CP)
	XLOC_006849	-1.049486635	cufflinks prediction (new CDS)
	Tb927.7.177	-1.03693738	expression site-associated gene (ESAG) protein,putative,expression site-associated gene 9 (ESAG9) protein,putative
	Tb927.6.510	-1.030149308	**GPEET2 procyclin precursor,PARP A-alpha,procyclin A-alpha,procyclic form specific polypeptide A-alpha precursor**
	Tb927.9.7370	-1.015957865	expression site-associated gene 9 (ESAG9) protein, putative
	XLOC_006811	1.031993577	cufflinks prediction (new CDS)
	Tb927.7.6530	1.231148733	variant surface glycoprotein (VSG, pseudogene), putative,variant surface glycoprotein, degenerate
	Tb927.2.1350	1.327252269	**retrotransposon hotspot protein (RHS, pseudogene), putative,retrotransposon hotspot protein 1 (RHS1), interrupted**
	Tb927.7.6510	1.421212112	variant surface glycoprotein (VSG, pseudogene), putative,variant surface glycoprotein, degenerate
	Tb927.7.6520	1.84048325	variant surface glycoprotein (VSG, pseudogene), putative,variant surface glycoprotein, frameshift
	Tb11.15.0035b	1.872283169	**hypothetical protein, conserved, degenerate**
	XLOC_004041	2.248717612	cufflinks prediction (new CDS)
up-regulated in dKO	Tb927.7.6500	3.579198515	variant surface glycoprotein (VSG), putative

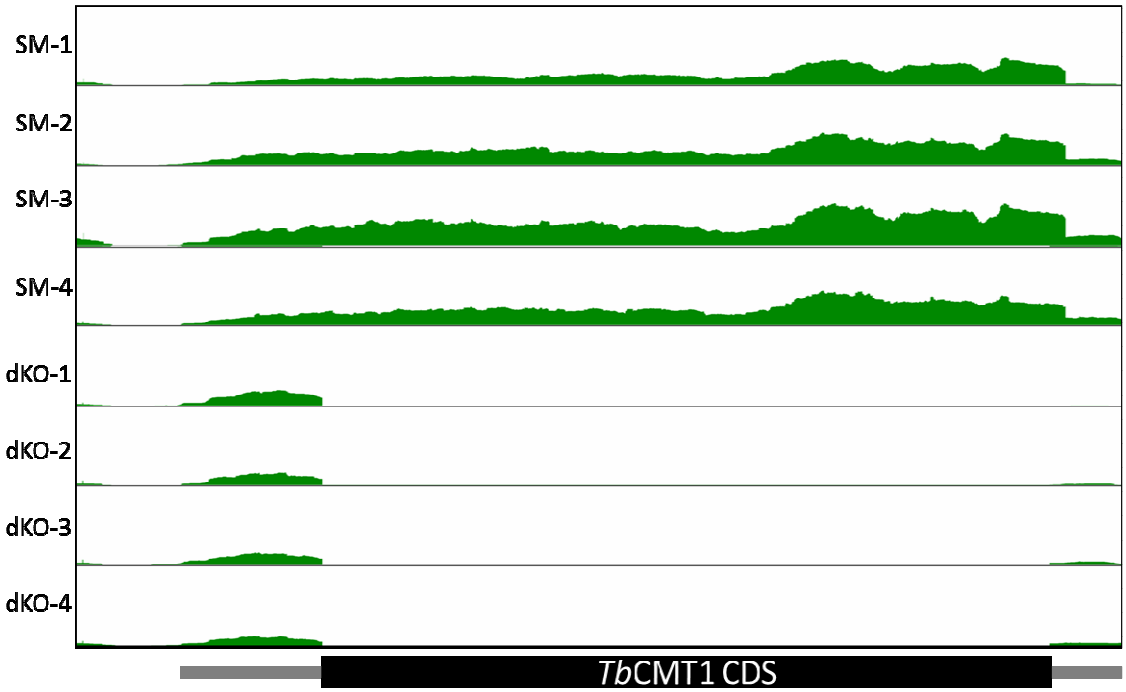


Figure 3.11: RNA-seq coverage plots for *TbCMT1* mRNAs. The number of sequence reads on the y-axis is shown along the length of the TbCMT1 gene in the four biological replicates of the SM (SM1-4) and *Tb*CMT1 null (dKO) *T. brucei* cell lines (dKO1-4). Schematic of the transcript shows the coding sequence (black) and UTRs (grey). Plot generated by Michele Tinti, MAJF lab.

3.5 Essentiality of the bifunctional guanylyltransferase-methyltransferase *Tb*CGM1

Despite several attempts, it was not possible to generate conditional null mutants of *Tb*CGM1 (Figure 3.12, panel A). Although cells recovered after electroporation, initial attempts to replace the first allele with the PAC resistance cassette failed due to insertion of the cassette into incorrect loci. In case of haploinsufficiency, the *Tb*CGM1 ectopic copy was electroporated into the cells before replacement of the first allele. Indeed, with the ectopic copy present, the first allele was successfully replaced. However, the tetracycline regulation of the ectopic copy was lost and as a result it was expressed constitutively (data not shown). This further suggested the possibility that bloodstream form *T. brucei* were haploinsufficient for *Tb*CGM1. An alternative ectopic copy expression vector (pLew100 v5) with higher tet-induced expression was introduced to maintain higher levels of *Tb*CGM1 during allele replacement (Figure 3.12, panel B). Western blot analysis of recovered cells showed tet-induced expression of the HA-tagged protein with no detectable expression in un-induced cells (Figure 3.13). Nonetheless, replacement of the first allele with the PAC resistance cassette was unsuccessful and once again the cassette was introduced into the wrong locus. Finally, the first allele was replaced with the HYG resistance cassette, but when the *TbCGM1*$^{-/+}$, *TbCGM1*^{Ti} cells were cultured in non-permissive conditions the growth rate was not impaired, disproving the previous hypothesis that the cells are haploinsufficient for *Tb*CGM1. Attempts to replace the second allele with the PAC resistance cassette were yet again met with misincorporation of the construct (data not shown).

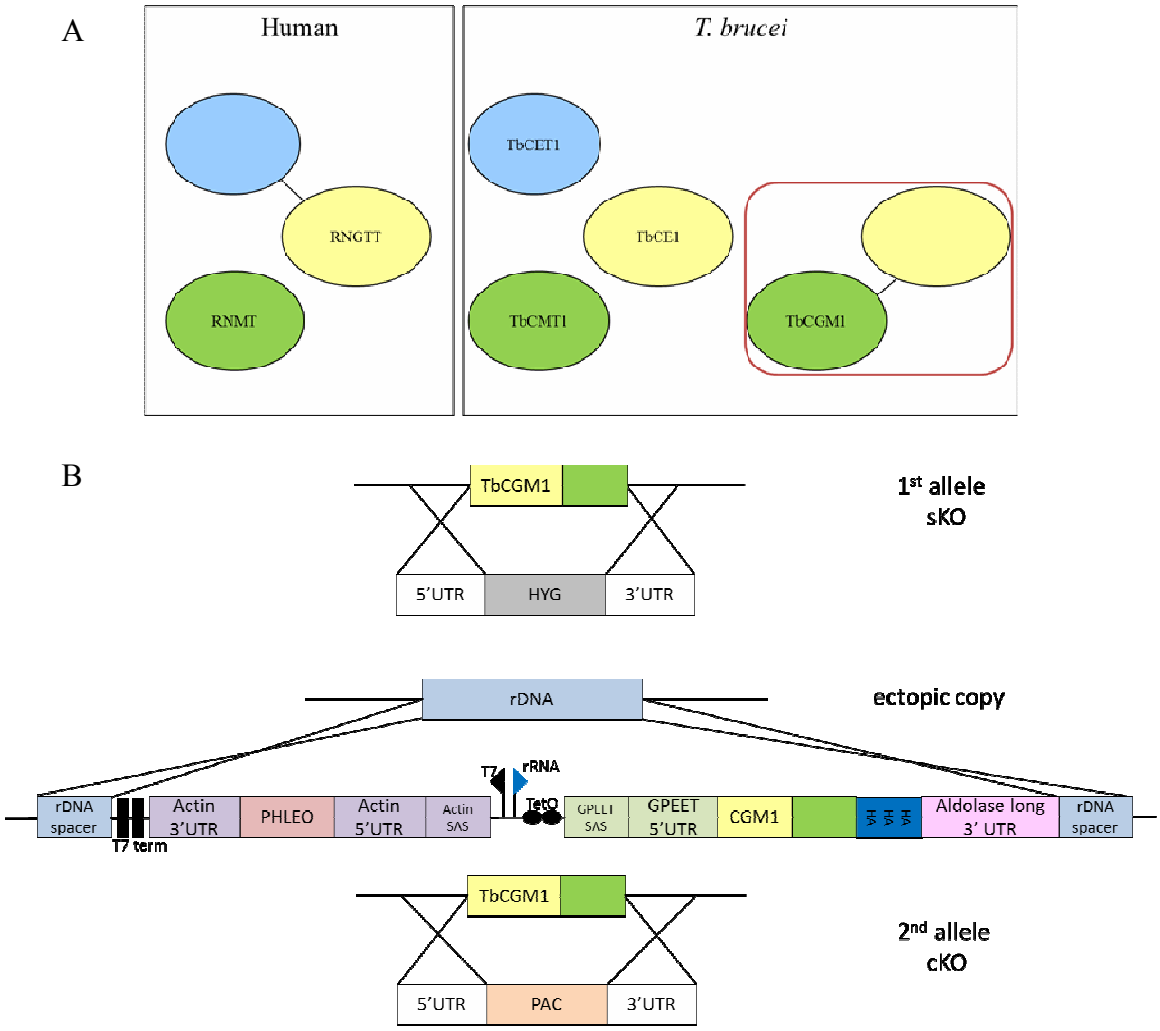


Figure 3.12: Bloodstream form *T. brucei TbCGM1* gene replacement strategy (A) Cartoon representing the arrangement of human and *T. brucei* mRNA capping enzymes, red box outlines bifunctional enzyme*Tb*CGM1. (B) The first allele of *TbCGM1*was replaced with a HYG resistance cassette. An ectopic copy of a tetracycline-inducible version of *Tb*CGM1 with three C-terminal HA tags was introduced into the rDNA locus using a rRNA promoter under the control of two tet-operator (TetO) sequences. Selection was via a phleomycin (PHLEO) resistance marker driven by a T7 promoter and suppressed by two T7 terminators. Activation of the ectopic copy with tetracycline should allow replacement of the second allele, regardless of essentiality, with a PAC resistance cassette.

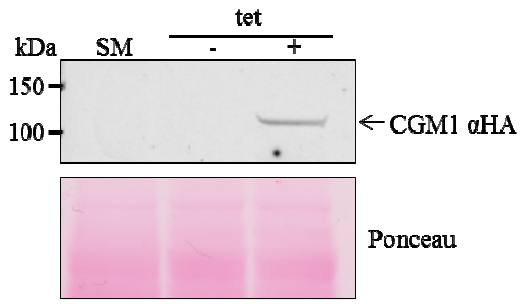


Figure 3.13: Western blot analysis of bloodstream form *T. brucei* expressing a3 HA-tagged *Tb*CGM1 ectopic copy. Cell lysate ($1x10^7$ cells/lane) of SM, and cells expressing *Tb*CGM1-3HA with the pLew100v5 vector grown in permissive (+ tet) and non-permissive (- tet) conditions were resolved by SDS-PAGE, blotted on nitrocellulose, and probed with anti-HA antibody. The membrane was stained with Ponceau before blocking to verify equal protein loading in each well.

3.6 Essentiality of *Tb*CGM1 by RNAi knockdown

Following the inconclusiveness of essentiality by gene replacement, inducible RNA interference (RNAi) was used in bloodstream form *T. brucei*. A 535 bp fragment of the *TbCGM1* ORF was cloned into a pRPa^{iSL} stem-loop RNAi vector. The resulting construct was electroporated into 2T1 *T. brucei* cells to generate a cell line expressing a tetracycline inducible double-stranded RNA (dsRNA) targeting *TbCGM1*. Correct integration of the construct into the rRNA spacer was assayed first by hygromycin selection followed by a check for puromycin sensitivity. Activation of the dsRNA targeting *TbCGM1* resulted in 67 ± 17% knockdown of *Tb*CGM1 mRNA after 24 h (Figure 3.14, panel A). *Tb*CGM1 knockdown resulted in reduction of the growth rate after 24 h that led to cell death after 48 h, demonstrating the essentiality of *Tb*CGM1 in culture (Figure 3.14, panel B).

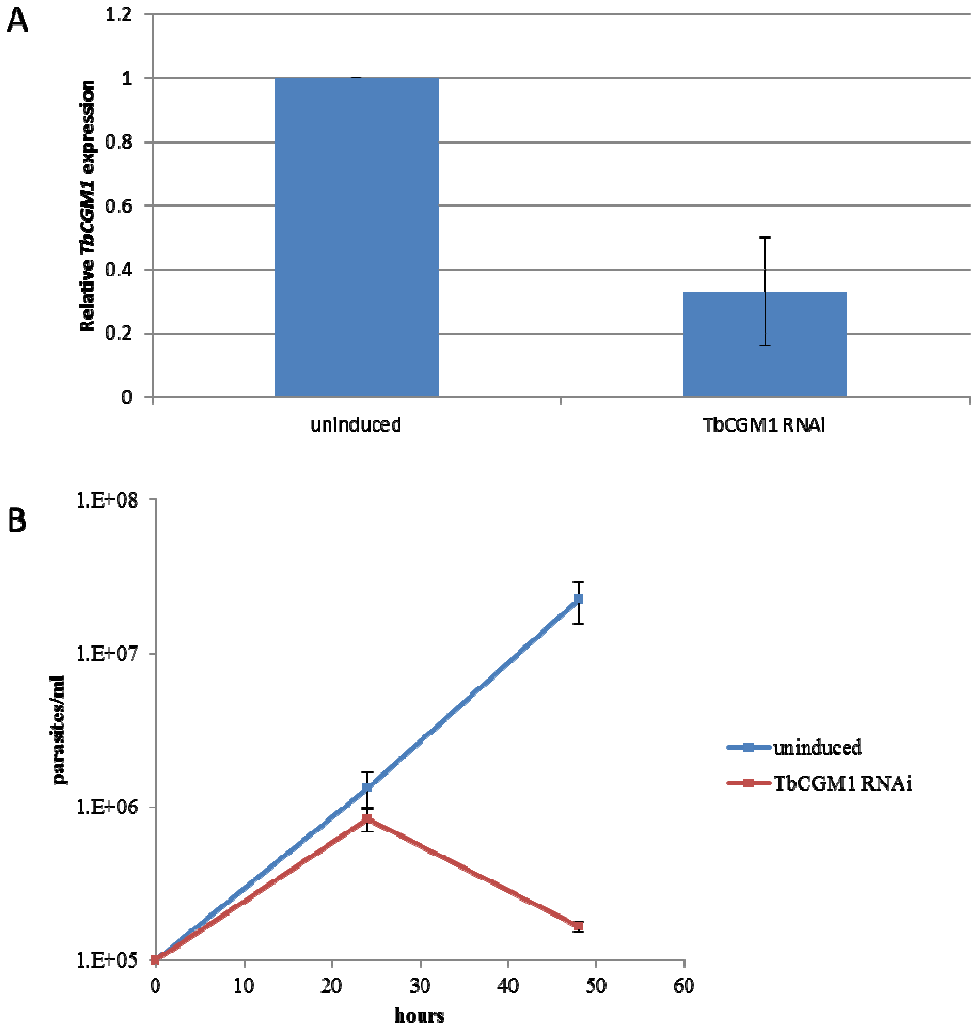


Figure 3.14: *Tb*CGM1 is essential for bloodstream form cell growth *in vitro*. (A) Total RNA from cells that were uninduced or tetracycline induced to knockdown *Tb*CGM1 for 24 hours was purified and analysed by qRT-PCR to analyse the downregulation of *Tb*CGM1 transcript. Three biological replicates were analysed; the mean ± standard deviations are shown. Results were normalized to the constitutive gene: telomerase reverse transcriptase (TERT) (Brenndörfer and Boshart, 2010). (B) Growth curve of cells that were uninduced or tetracycline induced to knockdown *Tb*CGM1with gene specific RNAi. Cells were grown in parallel; cultures were counted every 24 h and diluted to $1x10^5$ cells/ml every day. Induced cells were treated with 1µg/ml tetracycline every day. Three biological replicates were counted in triplicate; the mean ± standard deviations are shown.

RESULTS – CHAPTER II

RESULTS – CHAPTER II

4. Structural investigation of the *T. brucei* mRNA 5′ cap

4.1 Theoretical products of trifluoroacetic acid hydrolysis

The mRNA 5′ cap is required for to transport the transcript from the nucleus to the cytoplasm, maintain mRNA stability and allow for the efficient translation the transcript to protein. To investigate the possibility that some of the mRNA 5′ cap nucleosides in *T. brucei* are structurally different from conventional 7-methylguanosine, a method was developed to isolate the cap from endogenous mRNA and analyse the resulting structure by mass spectrometry. The possibility of novel mRNA cap structure(s) was suggested by the existence of an additional guanylyltransferase/methyltransferase (*Tb*CGM1) found in kinetoplastids but is absent in other eukaryotes. The structure of the typical eukaryotic mRNA methyl cap is an inverted guanosine residue methylated at the nitrogen 7 position. The 7-methylguanosine cap is joined to the first transcribed nucleotide by a 5′-5′ triphosphate linkage (Figure 4.1). This linkage is different from the 5′-3′phosphodiester bonds which join the rest of the nucleotides in an RNA molecule. To explore the chemical properties of the cap, we used the commercially available analogue m7GpppG. This analogue mimics the 5′end of an mRNA transcript with 7-methylguanosine connected to a guanosine (representing the first nucleotide) via a triphosphate bridge (Figure 4.2). In *T. brucei*, the first nucleotide of the splice leader sequence is a tri-methylated adenosine (Bangs et al., 1992). However, for the purpose of developing the mRNA cap isolation method, the guanosine in the analogue was deemed to be acceptable.

Trifluororacetic acid (TFA) is an acid that lends itself to the hydrolysis of biopolymers and its volatility expedites removal of the reagent before the subsequent steps of any protocol. The first step of the method was to use mild acid (TFA) hydrolysis to exploit

the extreme acid lability of pyrophosphate bonds (relative to 5′ or 3′phosphodiesterbonds). In a cell, polyphosphate (including pyrophosphate) bonds are hydrolysed by enzymes and in the absence of enzymatic catalysis their hydrolysis is very slow, except in under acidic conditions (Huebner and Milburn, 1980; Van Wazer et al., 1955). We made use of this principle to differentiate between the triphosphate bond linking the 5′ cap to the transcript, and all of the other bonds connecting individual nucleotides in a transcript (Figure 4.1). The theoretical products of the TFA hydrolysis of the methyl cap analogue m7GpppG are shown in Figure 4.3. A time course of the TFA hydrolysis of m7GpppG was performed and the products of the reaction were resolved by thin layer chromatography (TLC). A range of standards were used to compare the results of hydrolysis with the theoretical reaction products (Figure 4.4). Unfortunately, the products generated by TFA hydrolysis of m7GpppG were more complex than expected and consequently all further analysis was carried out using negative ion electrospray mass spectrometry.

4.2 Optimization of TFA hydrolysis

A mixture of the potential products of TFA hydrolysis was resolved by mass spectrometry (Figure 4.5). The differences in mass between the compounds yielded clear distinction between the standards. Next, the hypothesis that TFA will predominantly cleave adjacent phosphate bonds and not disturb the ribose sugar-base portion of the cap structure was investigated by hydrolysing components of the cap analogue as well as theoretical hydrolysis products (Figure 4.6). As predicted, TFA does not hydrolyse guanosine (Figure 4.6, panel A), or guanosine monophosphate (GMP) (Figure 4.6, panel B).

-

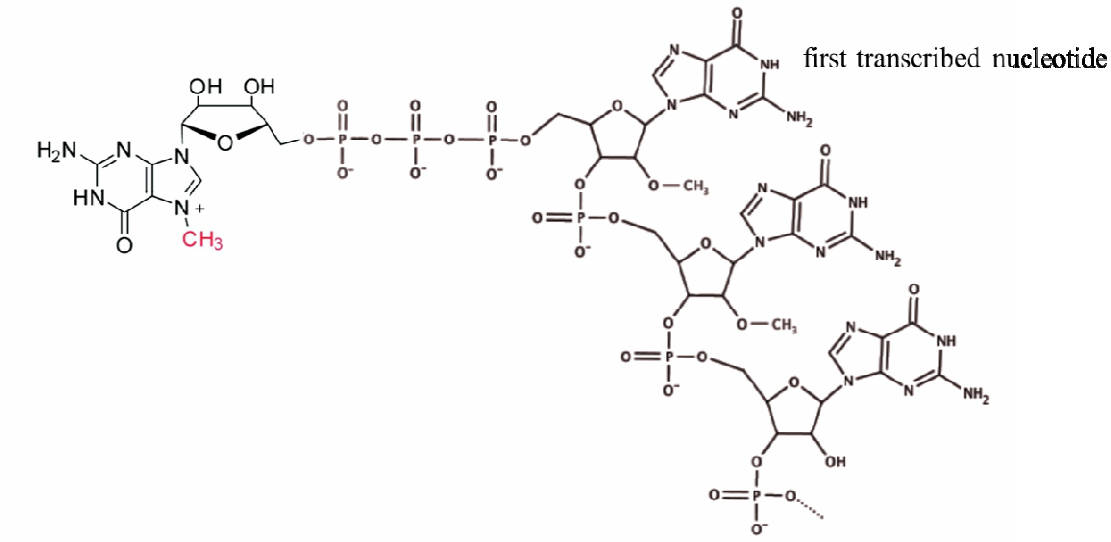


Figure 4.1: Structure of mRNA 5' cap. The mRNA 5'cap structure is composed of a 7-methylguanosine linked to the body of the mRNA transcript through a 5'-5' triphosphate bridge. The N7 methylation of the guanosine is outlined in red.

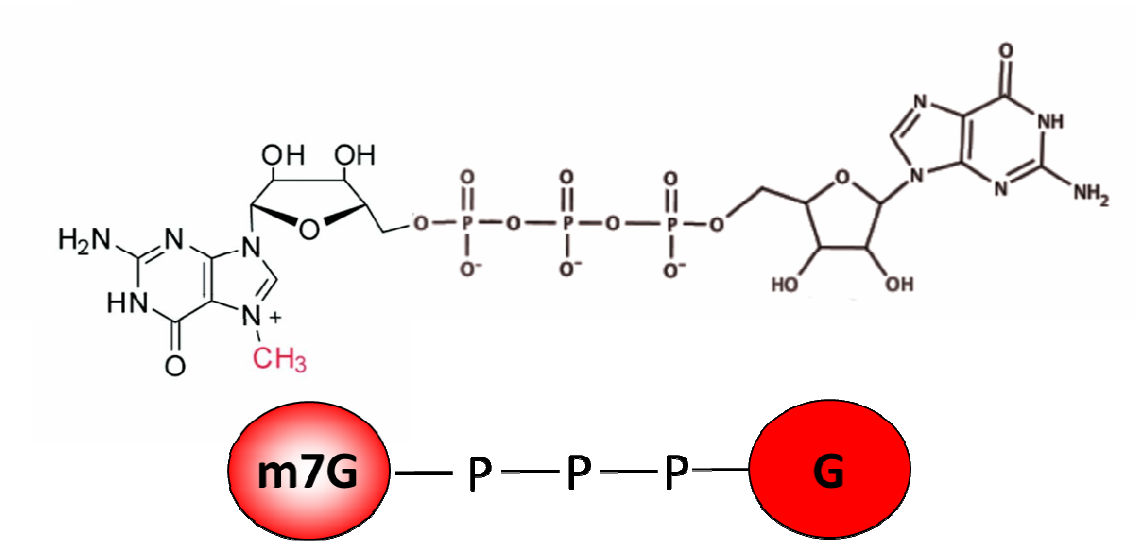


Figure 4.2: mRNA cap structure analogue. 7-methylguanosine linked to a guanosine through a 5'-5' triphosphate bridge. = 7-methylguanosine, = guanosine, P = phosphate group.

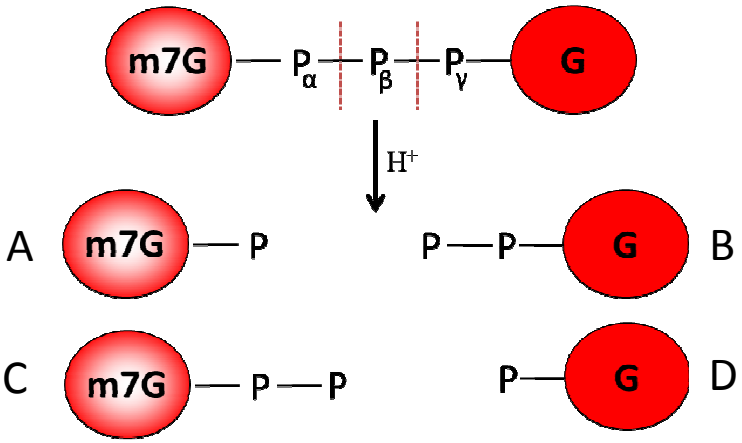


Figure 4.3: Theoretical products of TFA hydrolysed cap structure analogue, m7GpppG. Predicted sites for hydrolysis are designated by the dotted red lines. Cleavage between the α and β phosphate groups would yield products A (m7GMP) and B (GDP); cleavage between the β and γ phosphate groups would yield products C (m7GDP) and D (GMP); and cleavage of both possible sites would yield products A and C.

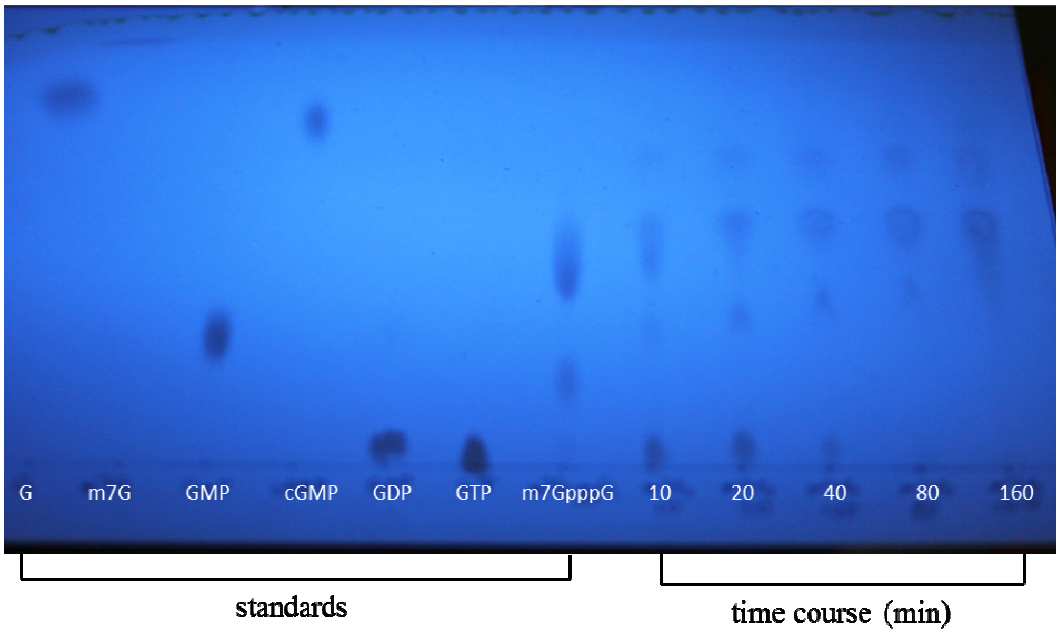


Figure 4.4: TFA hydrolysis time course of cap structure analogue, m7GpppG. Six standards: (guanosine (G), 7-methylguanosine (m7G), guanosine monophosphate (GMP), cyclic GMP (cGMP), guanosine diphosphate (GDP), and guanosine triphosphate (GTP) and the starting material: m7GpppG were resolved by TLC alongside the products of TFA hydrolysis with increased time.

7-methylguanosine could not be analysed because the methylation at the N7 position confers a positive charge to the cap structure and our analysis was carried out using the negative ion mode of the mass spectrometer. However, in the subsequent experiments described below, the cap structure was analysed in the form m7GDP which was readily detected. When hydrolysed with TFA, the pyrophosphate bond in guanosine diphosphate (GDP) yielded GMP (Figure 4.6, panel C) and guanosine triphosphate (GTP) yielded GDP and GMP as anticipated (Figure 4.6, panel D). The hydrolysis of GDP and GTP also generates the degradation product ribose 5-diphosphate but the amount of degradation is relatively small and should not interfere with the methodology (Figure 4.6, panels C and D). Taken together these results confirm TFA primarily hydrolyses pyrophosphate bonds and is therefore a reasonable choice of mild acid to release the majority of the cap while leaving most of the RNA intact.

4.3 TFA hydrolysis of cap analogue m7GpppG

Hydrolysis conditions were optimized to obtain the greatest yield of m7GDP (Figure 4.3, "product C"). Following confirmation that the starting material (m7GpppG) was pure (Figure 4.7, panel A), as a first approximation, the cap analogue was hydrolysed for 15 min (Figure 4.7, panel B). Evidence of both the starting material and desired reaction product (m7GDP) indicated 15 min is in the range of optimal hydrolysis. Next, a TFA hydrolysis time course was performed where time points were collected between 0 and 20 min (Figure 4.8). The 0 min time point served as a positive control for the stability of the starting material in the absence of acid (Figure 4.8, panel A). m7GDP was undetectable before 5 min of hydrolysis (Figure 4.8, panel B), very little was detected before 15 min (Figure 4.8, panels C and D), and the greatest yield of m7G was found after 20 min of TFA hydrolysis (Figure 4.8, panel F). More than 20 min of hydrolysis resulted in a greater proportion of the ribose-5-diphosphate degradation product than m7GDP (data not shown).

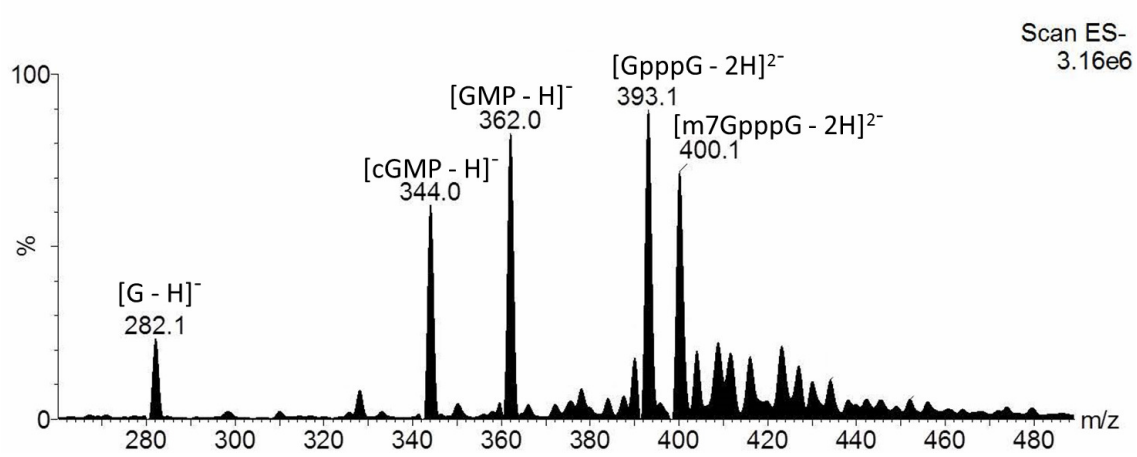
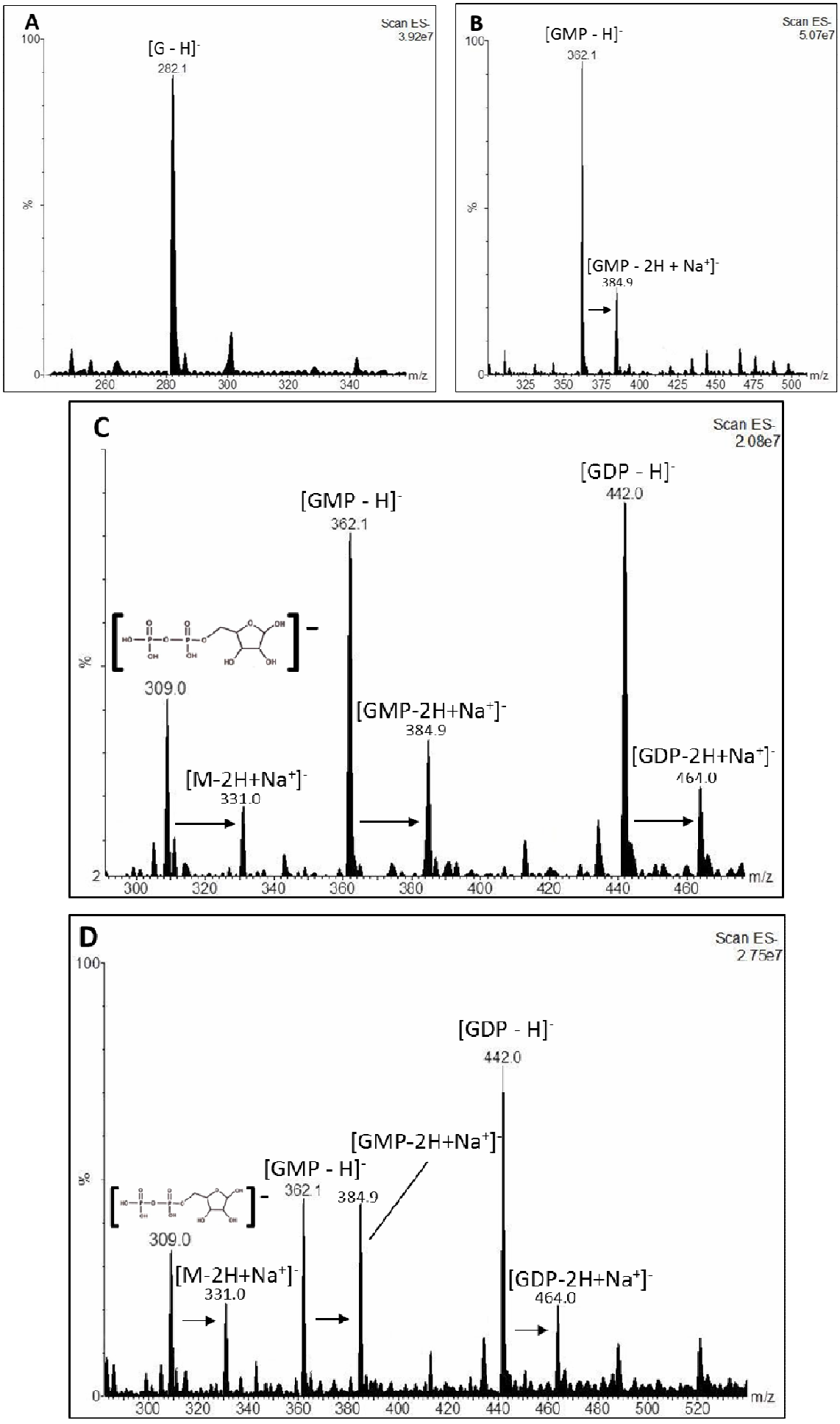


Figure 4.5: Chemical standards of theoretical hydrolysis products. Representative negative ion ES-MS chromatogram of a mixture of cap analogues (GpppG and m7GpppG) and the potential products of TFA hydrolysis (G, cGMP, and GMP).

Figure 4.6: *(next page)* TFA hydrolysis of theoretical reaction products. Representative negative ion ES-MS chromatograms of standards of G, GMP, GDP, and GTP hydrolysed with TFA. (A/B) The majority of G and GMP were not hydrolysed as seen by the $[M-H]^-$ ions at *m/z* 282 and *m/z* 362, respectively. (C) GDP observed as the $[M-H]^-$ ion at *m/z* 442 is hydrolysed into GMP seen by the $[M-H]^-$ ion at *m/z* 362 and the degradation product ribose diphosphate at *m/z* 309. (D) GTP is completely hydrolysed into GDP and GMP detected as the $[M-H]^-$ ions at *m/z* 442 and 362, respectively. The degradation product ribose diphosphate was also observed at *m/z* 309. Data collected on Ultima.The presence of sodium adduct ions $[M-2H+Na^+]^-$ are indicated.


A
Scan ES-
3.92e7
[G - H]-
282.1
100
%
0
260
280
300
320
340
m/z
B
Scan ES-
5.07e7
[GMP - H]-
362.1
[GMP - 2H + Na+]-
384.9
100
%
0
325
350
375
400
425
450
475
500
m/z
C
Scan ES-
2.08e7
[GDP - H]-
442.0
[GMP - H]-
362.1
309.0
[GMP-2H+Na+]-
384.9
[M-2H+Na+]-
331.0
[GDP-2H+Na+]-
464.0
%
2
300
320
340
360
380
400
420
440
460
m/z
D
Scan ES-
2.75e7
[GDP - H]-
442.0
[GMP-2H+Na+]-
[GMP - H]-
362.1
384.9
309.0
[M-2H+Na+]-
331.0
[GDP-2H+Na+]-
464.0
100
%
0
300
320
340
360
380
400
420
440
460
480
500
520
m/z

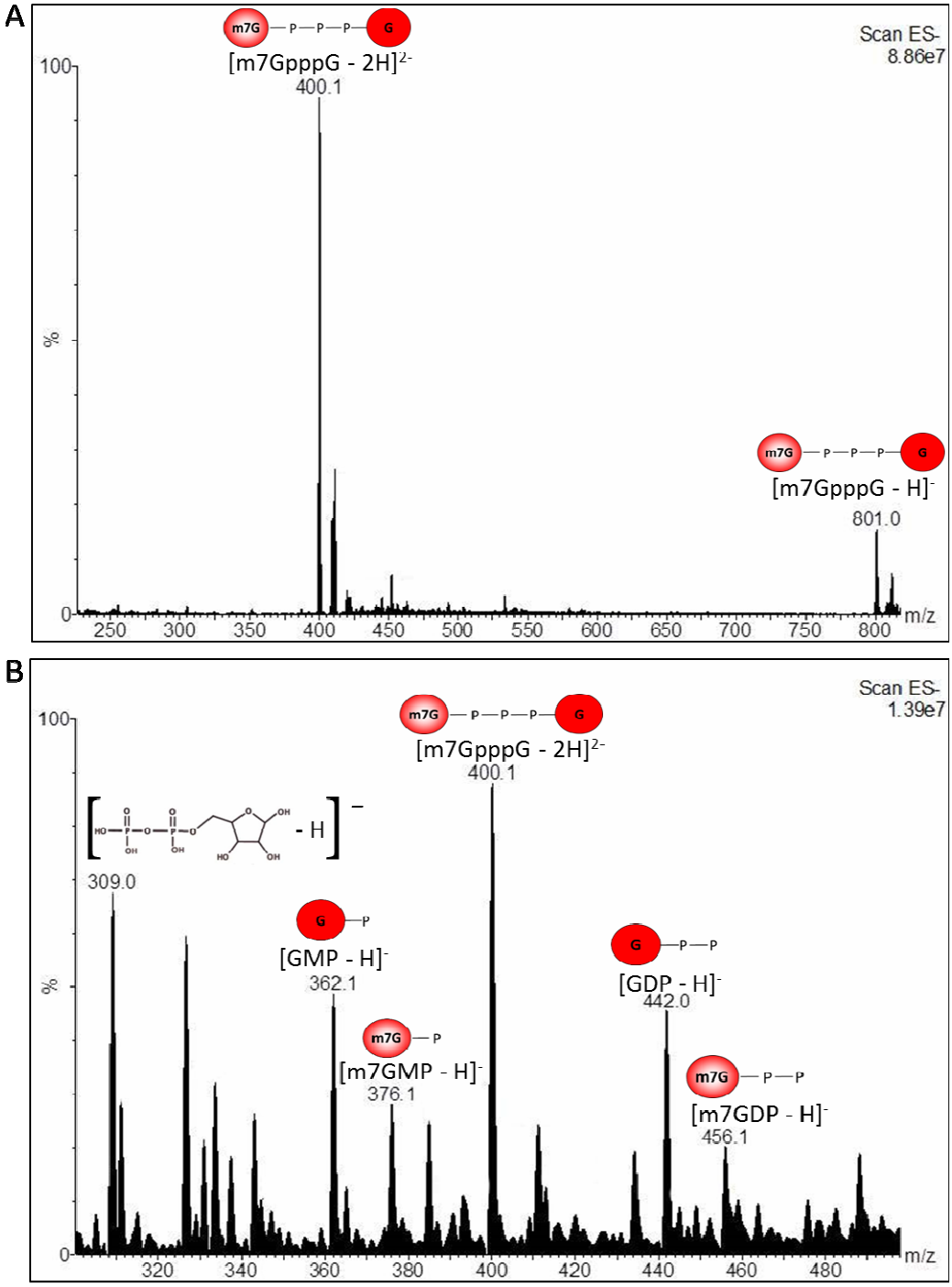


Figure 4.7: TFA hydrolysis of m7GpppG. (A) Representative ES-MS chromatogram of the starting material, cap analogue m7GpppG, was observed as the major $[M-2H]^{2-}$ ion at *m/z* 400 and as the $[M-H]^{-}$ ion at *m/z* 801. (B) TFA hydrolysis of m7GpppG shows non-hydrolysed starting material at *m/z* 400. Two diphosphate containing products are detected: the GDP ion at *m/z* 442 and the m7GDP ion at *m/z* 456. Side products GMP and m7GMP were detected at *m/z* 362 and *m/z* 376, respectively, as well as a ribose diphosphate degradation product at *m/z* 309. Data collected on Ultima.

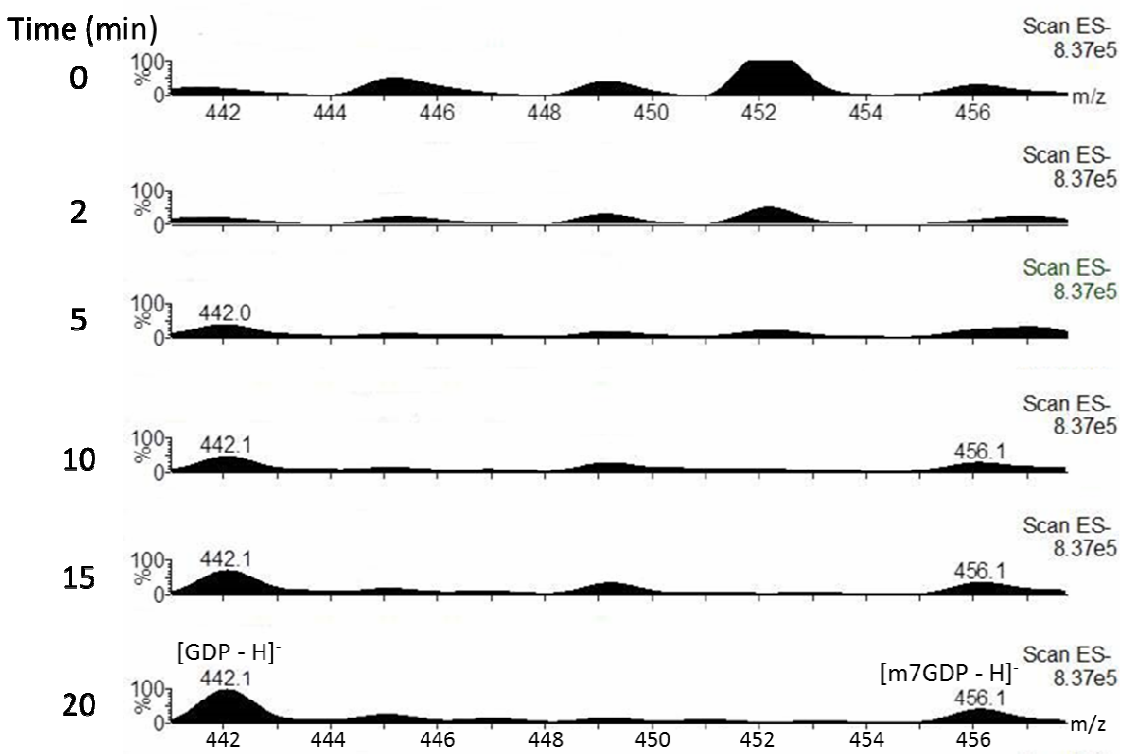


Figure 4.8: Time course of the TFA hydrolysis of m7GpppG. Time points were collected between 2 and 20 minutes. The two diphosphate containing products: the GDP ion at *m/z* 442 and the m7GDP ion at *m/z* 456 are detected with increasing intensity over the course of 20 m. Data collected on Ultima.

4.4 Precursor ion scanning mass spectrometry

Each nucleotide in an RNA molecule is connected to the adjacent nucleotide by a phosphate group. The exception to this arrangement is the 5´cap, which is connected to the first nucleotide by three phosphate groups as described above. This exception can be exploited by TFA hydrolysis because the cap is the only structure in mRNA that can be released with a pyrophosphate group. In precursor ion scanning mass spectrometry, the specific loss of an ion (in this case pyrophosphate) from the precursor molecule is monitored across the spectrum. Having optimized hydrolysis conditions that obtain the greatest yield of m7GDP, a precursor ion scanning method was established to detect nucleoside diphosphates (Figure 4.9).

To achieve this, negative ion mass spectrometry was performed on an authentic standard of GDP, which produces an intense $[M-H]^-$ ion at *m/z* 442 (Figure 4.10, panel A). This ion was then selected in MS1, collision induced dissociation (CID) was performed in the hexapole collision cell, and the product ions in MS2 were detected. As expected, one of these product ions arises from the elimination of the pyrophosphate group (Figure 4.10, panel B). The mass spectrometer was then set up in precursor ion scanning mode to search for precursors of the *m/z* 159 pyrophosphate ions. In this mode, the spectrum will only be of ions that produce an *m/z* 159 product ion. This highly selective filter enables the identification of pyrophosphate containing molecules amongst a complex mixture and therefore potential novel cap structures of any mass can be identified. The method will not be able to discriminate isobaric cap structures; however, differences in HPLC retention time and/or product ion MS/MS experiments of ions derived from mRNA, for example, *m/z* 456 ions should provide evidence of novel isobaric structures.

4.5 TFA hydrolysis coupled to precursor ion scanning mass spectrometry of m7GpppG

A trial precursor ion spectrum using GDP with the pyrophosphate precursor ion scanning mode yielded a strong *m/z* 442 precursor ion, as expected (Figure 4.10, panel C). TFA hydrolysis and the precursor ion scanning method were combined and tested on the cap analogue m7GpppG. Encouragingly, both predicted precursor ions were observed: the GDP ion at *m/z* 442 and the m7GDP ion at *m/z* 456 (Figure 4.11). Despite a predicted equal ratio of GDP to m7GDP after hydrolysis, the intensity of GDP in the spectrum appears approximately three times greater because the fixed positive charge on the N7 nitrogen of m7GDP reduces its ionizability in the negative ion mode of the mass spectrometer. Nonetheless, qualitatively, the m7GDP peak observed at *m/z* 456 indicated that this ion was effectively cleaved from the cap analogue by TFA hydrolysis and detected by precursor ion mass spectrometry.

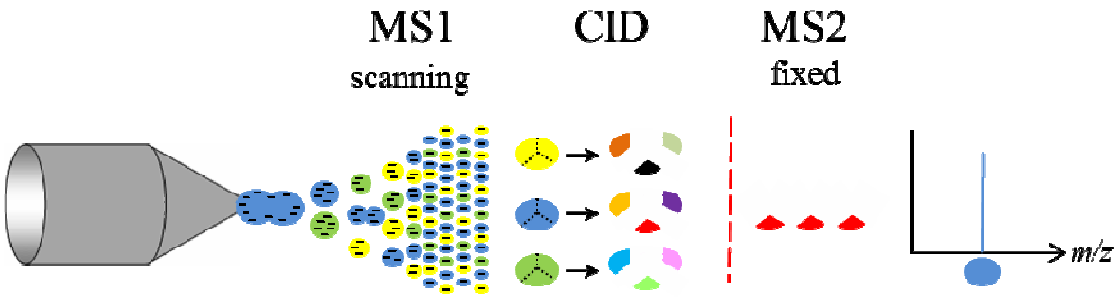
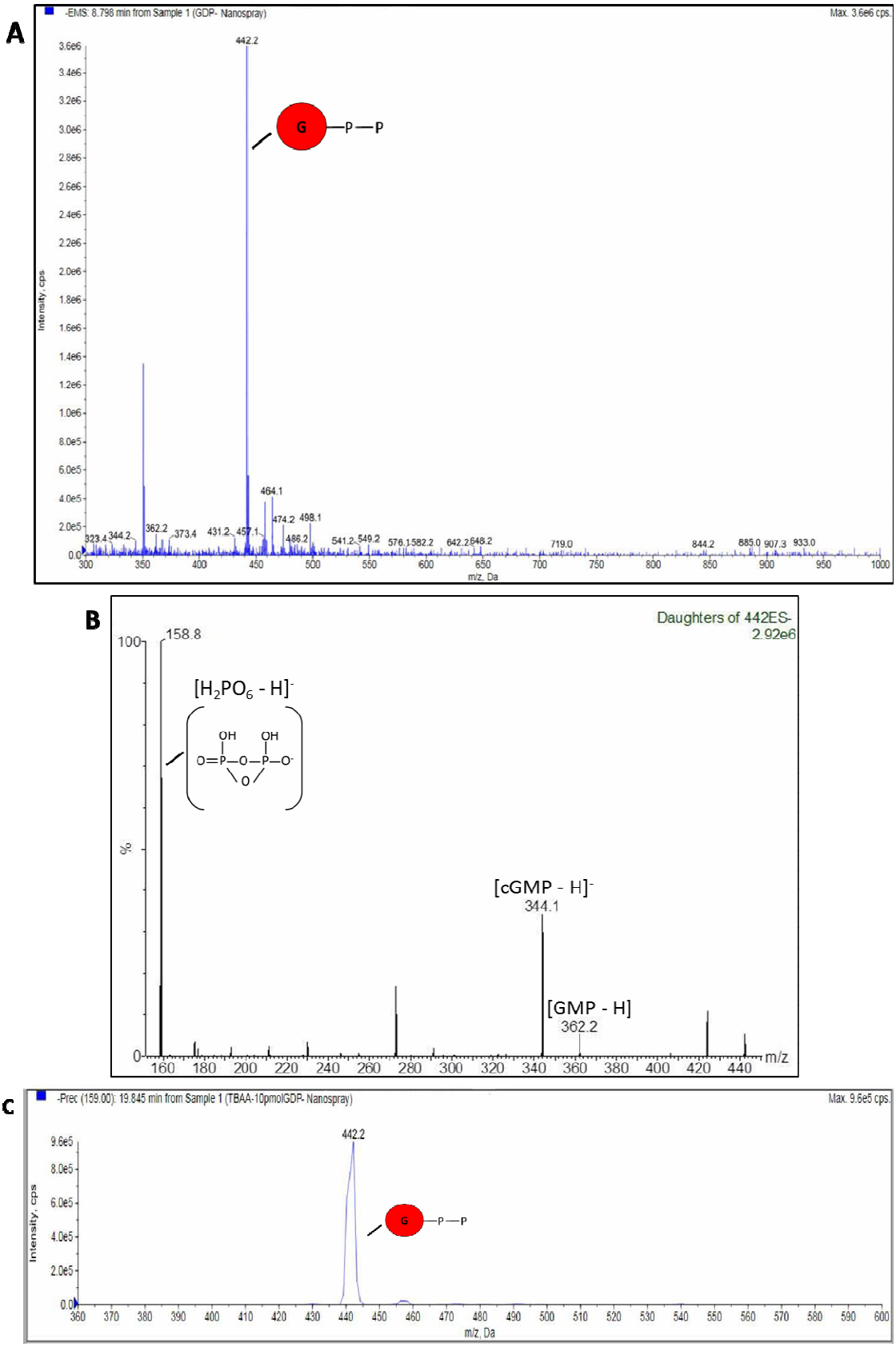


Figure 4.9: The principle of precursor ion scanning mass spectrometry. Ions pass through the first analyser (MS1) and then undergo collision-induced dissociation (CID) in the collision cell. The resulting fragments are analysed by the second analyser (MS2). Precursor ion scanning sets the MS2 analyser to select only one specific fragment ion (in this case the pyrophosphate ion *m/z* 159) to transmit to the detector. MS1 is then scanned to detect all of the precursor ions that generate this fragment.

Figure 4.10: *(next page)* Precursor ion scanning method. (A) MS1 spectrum of the GDP standard (G—P—P) confirmed by the *m/z* 442 ion consistent with $[M-H]^-$ species. Data collected on QTRAP 4000. (B) MS/MS spectrum of the product ions of *m/z* 442, including the pyrophosphate ion (*m/z* 159). Data collected on Ultima. (C) Precursor ion scan of *m/z* 159 produced an intense peak for predicted ion *m/z* 442 corresponding to GDP. Data collected on QTRAP 4000.


A
-EMS: 8.798 min from Sample 1 (GDP- Nanospray)
Max. 3.6e6 cps.
Intensity, cps
m/z, Da
442.2
G
P—P
464.1
B
Daughters of 442ES-
2.92e6
158.8
[H2PO6 - H]-
[cGMP - H]-
344.1
[GMP - H]
362.2
%
m/z
C
-Prec (159.00): 19.845 min from Sample 1 (TBAA-10pmolGDP- Nanospray)
Max. 9.6e5 cps.
Intensity, cps
m/z, Da
442.2

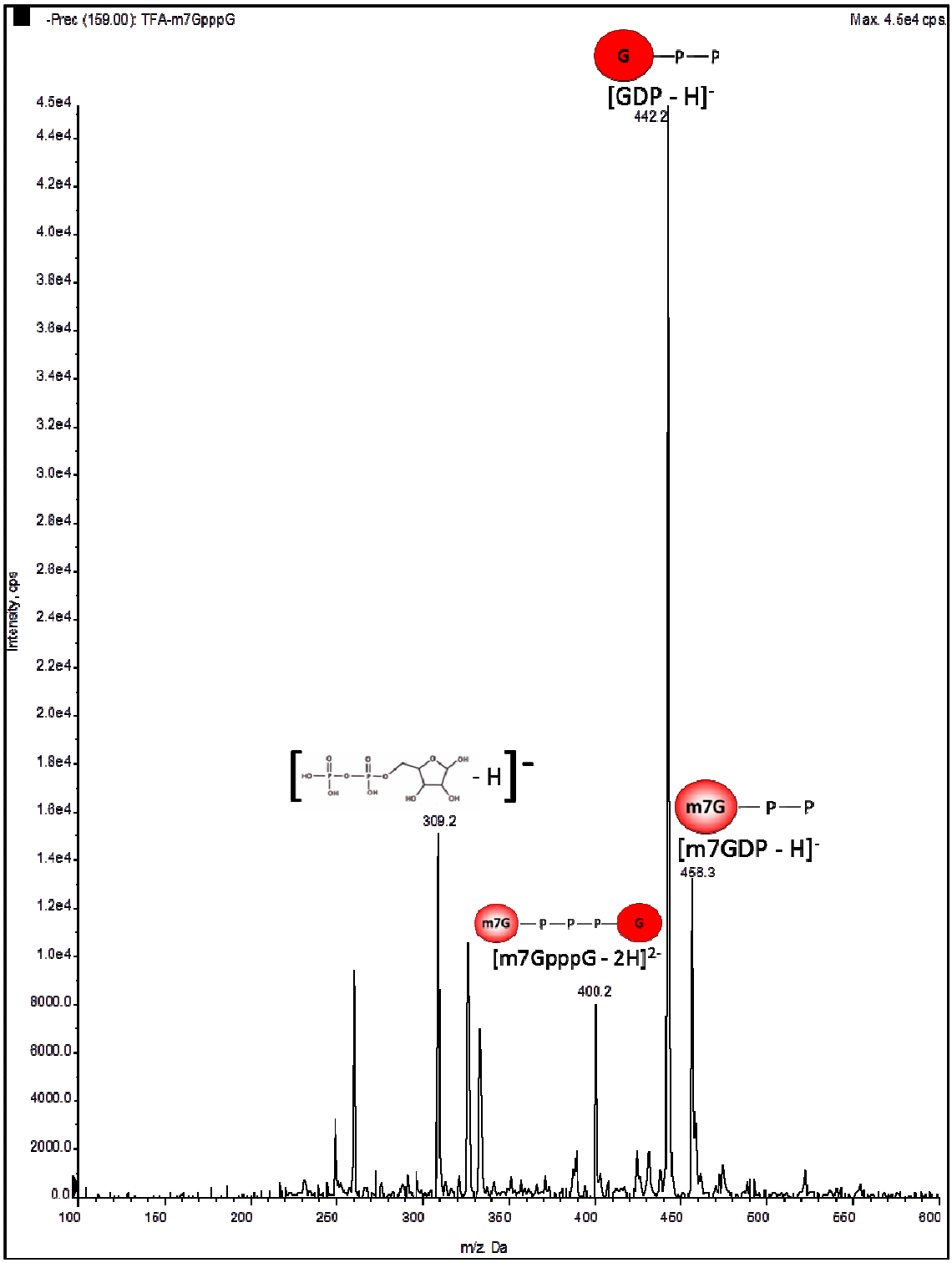


Figure 4.11: Precursor ion scan of TFA hydrolysed m7GpppG. Spectrum of the pyrophosphate ion (*m/z* 159) precursor ion scan produced intense peaks for the predicted pyrophosphate containing product ions: GDP ion at *m/z* 442 the m7GDP ion at *m/z* 456. Non-hydrolysed starting material (*m/z* 400) and ribose diphosphate (*m/z* 309) are also detected. Data collected on QTRAP 4000.

4.6 Mass spectrometer sensitivity optimization

The method was then ready to analyse *in vitro* transcribed RNA to gauge sensitivity with transcripts. Several preparations of *in vitro* transcribed RNA were prepared, but neither GDP nor m7GDP precursor ions could be detected (data not shown). The m7GDP ions cleaved off of the *in vitro* transcript may have been below the limit of detection of the mass spectrometer because the quantity of hydrolysed cap was less certain than in the experiments with the cap analogue. To address this, titrations of the GDP standard were used as a surrogate for the cap structure to optimize the sensitivity of the mass spectrometer. An intense peak of the GDP $[M-H]^-$ ion at *m/z* 442 was achieved at 10 pmol, a significant improvement on the initial sensitivity of 1 nmol (Figure 4.12). With this improved sensitivity, the method was tested on mammalian HEK-293 cell total RNA (data not shown). Although the precursor ion scanning method is a very selective filter, the complexity of cell culture derived RNA revealed that the signal to noise ratio in the spectrum was too low to observe the 5′ cap diphosphate ion m7GDP. Increasing the amount of input RNA would not improve the spectrum as the number of unidentifiable precursor ions would also increase proportionally. To purify the sample prior to hydrolysis and analysis, the mRNA was selectively enriched out of the mixture of tRNA and highly abundant rRNA.

4.7 Poly(A) mRNA binding buffer optimization

In addition to the 5′ cap, mRNA is characteristically polyadenylated on its 3′ end. This polyadenylated region can be used to efficiently recover mRNA from the other RNA molecules using oligo-dT cellulose. In order to use this technique in conjunction with mass spectrometry, it was necessary to identify a suitable substitute for the oligo-dT buffers because salts and detergents cannot be used in mass spectrometers and further purification may cause sample loss. Salts overload the system with charged salt ions and generate multiple salt adducts of the desired analyte.

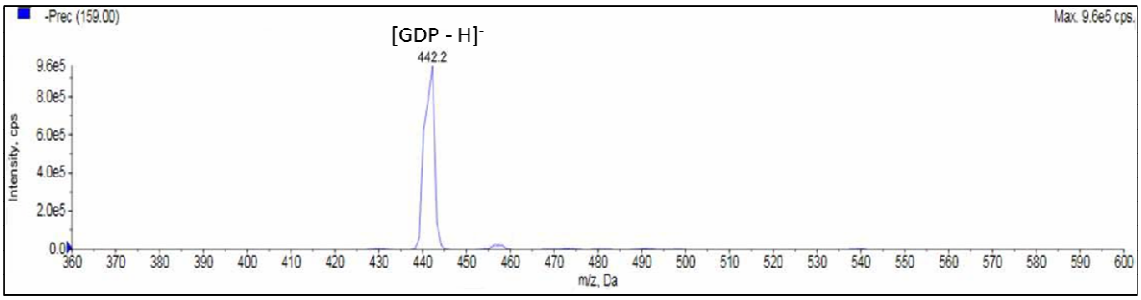


Figure 4.12: Precursor ion scan of GDP standard. Spectrum of the pyrophosphate ion (*m/z* 159) precursor ion scan of 10 pmol of pure GDP produced an intense peak for the predicted pyrophosphate containing GDP ion at *m/z* 442. Data collected on QTRAP 4000.

Therefore, components of the oligo-dT cellulose binding buffer were methodically removed and replaced with salts that can be better tolerated by the system. To explore the effects of binding buffer modification on mRNA extraction, total RNA from HEK-293 cells was enriched for mRNA using five different buffers. The positive control for effective mRNA purification was the standard binding buffer (Figure 4.13, panel A). Alone, the removal of the detergent sodium dodecyl sulphate (SDS) or the decrease in ethylenediaminetetraacetic acid (EDTA) concentration resulted in residual rRNAobserved in the three characteristic rRNA peaks(Figure 4.13, panels B and D). The removal of and SDS, Tris(hydroxymethyl)aminomethane (Tris) and substitution of sodium chloride (NaCl) with sodium acetate (NaOAc) reduced the mRNA binding capacity of the oligo-dT cellulose resulting in a lower yield, though the purified mRNA did not contain residual rRNA (Figure 4.13, panel C). Further alteration of the buffer by additionally decreasing the EDTA concentration proved unsuccessful with both a low yield and residual rRNA (Figure 4.13, panel E). Buffer number three was chosen for mRNA purification because it gave the purest yield and was incorporated into subsequent experiments (Figure 4.13, panel C).

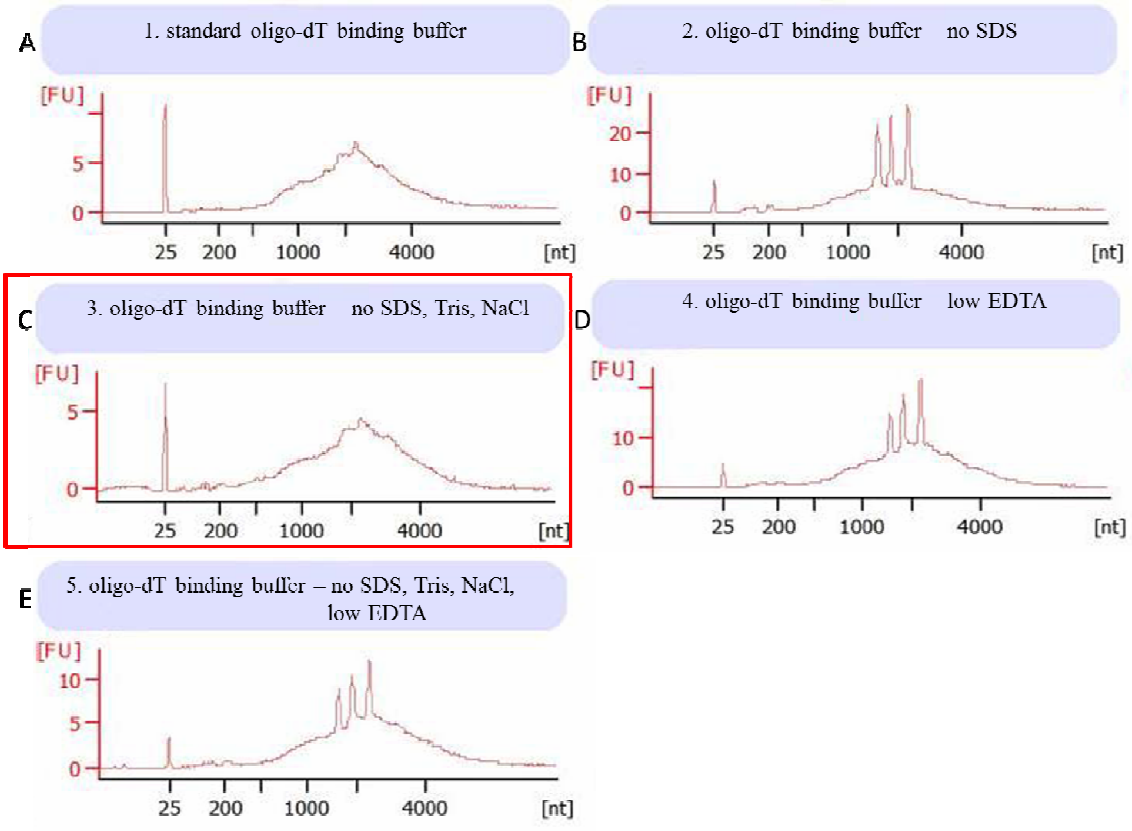


Figure 4.13: Modification of the oligo-dT binding buffer for mass spectrometry. Bioanalyzerelectropherograms of *T. brucei* mRNA purified using oligo-dT cellulose with variable binding buffers. (A) the standard binding buffer: 0.5M NaCl, 1mM EDTA, 10mM Tris, 0.05% SDS, (B) the standard buffer without SDS: 0.5M NaCl, 1mM EDTA, 10mM Tris, (C) the standard buffer without SDS, Tris, and NaCl: 0.5M NaOAc, 1mM EDTA, (D) the standard buffer with a lower concentration of EDTA: 0.5M NaCl, 1µM EDTA, 10mM Tris, 0.05% SDS, (E) the standard buffer without SDS, Tris, NaCl, and with a lower concentration of EDTA : 0.5M NaOAc, 1µM EDTA. Peak at 25 nt indicates fluorescence standard. nt= nucleotides, FU= Fluorescence Units

4.8 TFA hydrolysis method trial with HEK-293 cell mRNA

Combing these improvements, the method was re-designed (Figure 4.14). HEK-293 cells were used to assess the complete method from the culture flask to the precursor ion scan (Figure 4.15). An extracted ion chromatogram (XIC) filters the spectra to display a particular precursor ion. The XIC for the m7GDP ion at *m/z* 456 showed a peak at 15.89 m (Figure 4.15, panels A and B). This peak had only marginally greater intensity than the adjacent unknown peak at 21.16 m. However, the elution time of the m7GDP ion corresponded to the elution time of the m7GDP ion generated by hydrolysis of the cap analogue. The XIC of the unmethylated cap, GDP ion at *m/z* 442 produced an intense peak at 20.35 m (Figure 4.15, panels C and D). Ions tentatively corresponding to ADP at *m/z* 426 and CDP at *m/z* 402 were also detected. However, their identity has not been confirmed with an authentic standard. The intensity of the GDP peak was unexpected as the majority of the mRNA cap is thought to be a methylated (Rottman et al., 1974). Taken together, there is evidence that at least a proportion the mammalian mRNA cap is 7-methylguanosine.

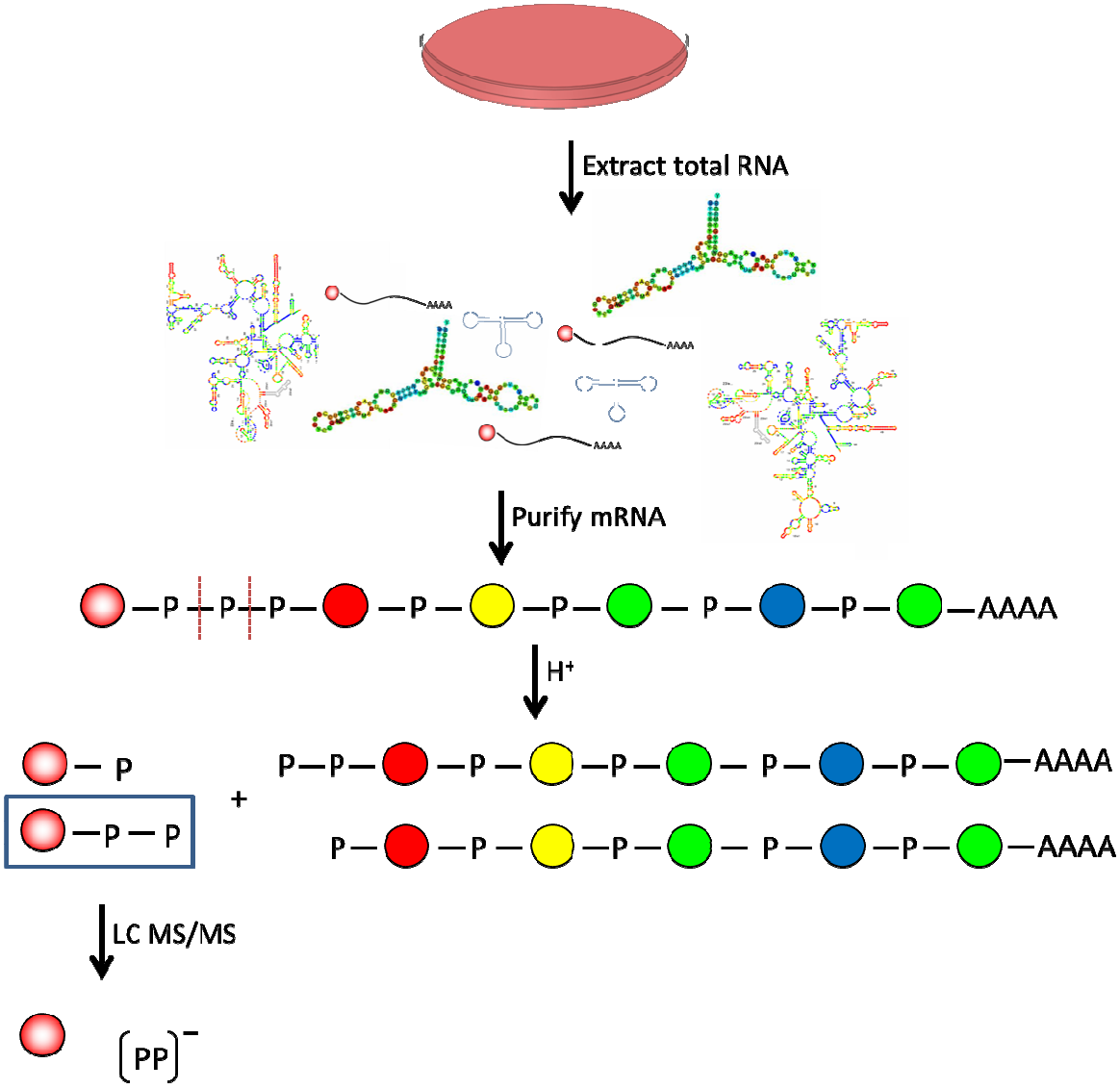
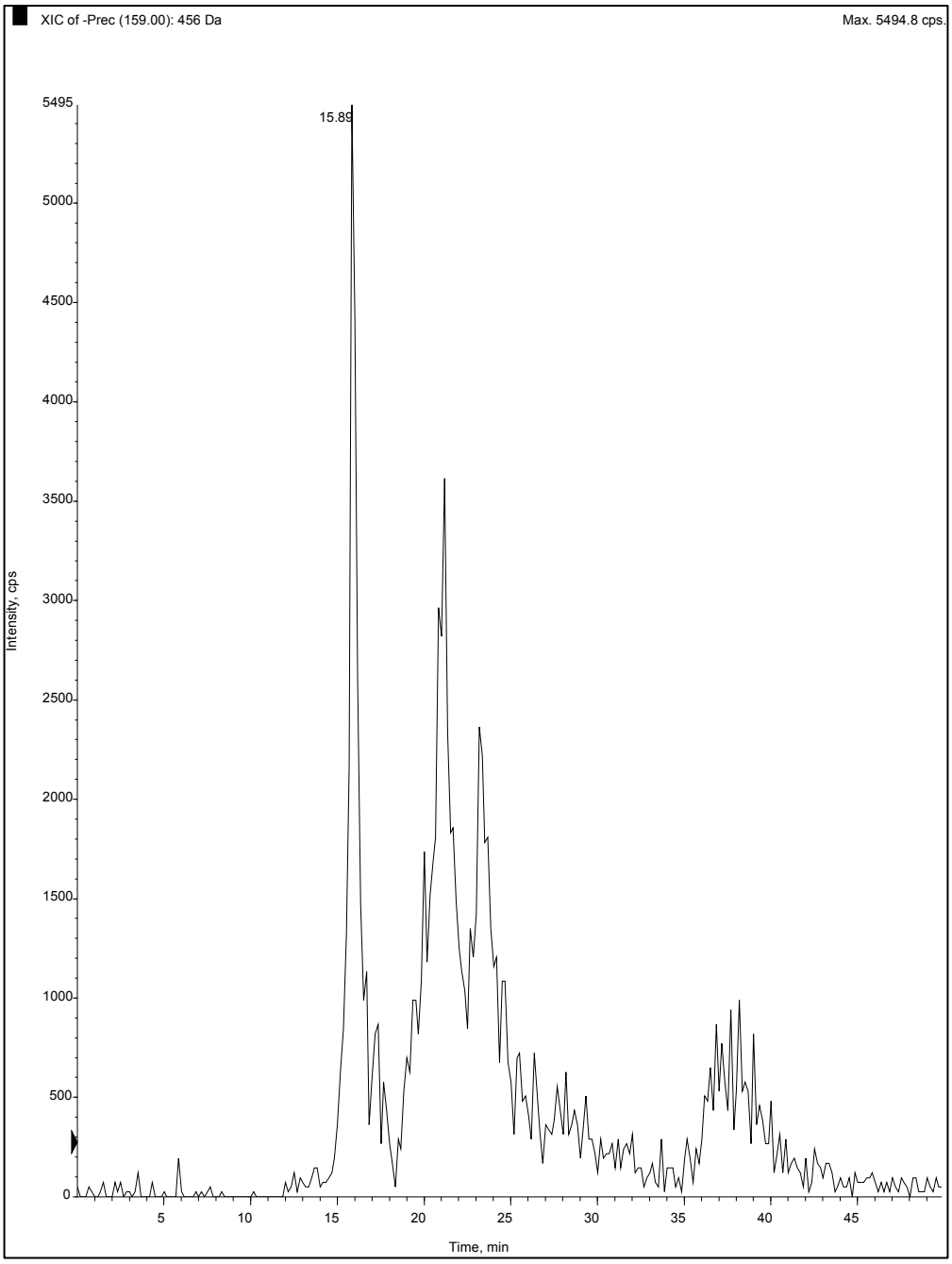
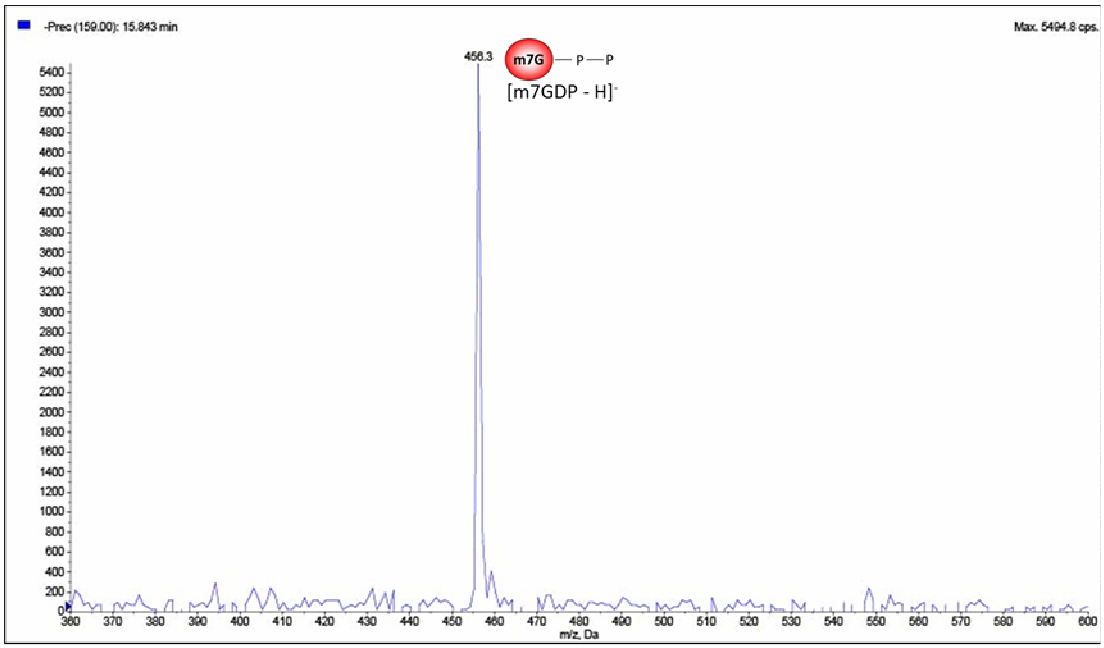


Figure 4.14: Schematic of TFA hydrolysis of mRNA. Red dashed lines indicate predicted cleavage of the molecule upon treatment with TFA and boxed product indicates m7GDP ion. Following hydrolysis, precursor ion mass spectrometry method was developed to identify the 5'cap structure which yielded a pyrophosphate molecule upon collision induced dissociation.

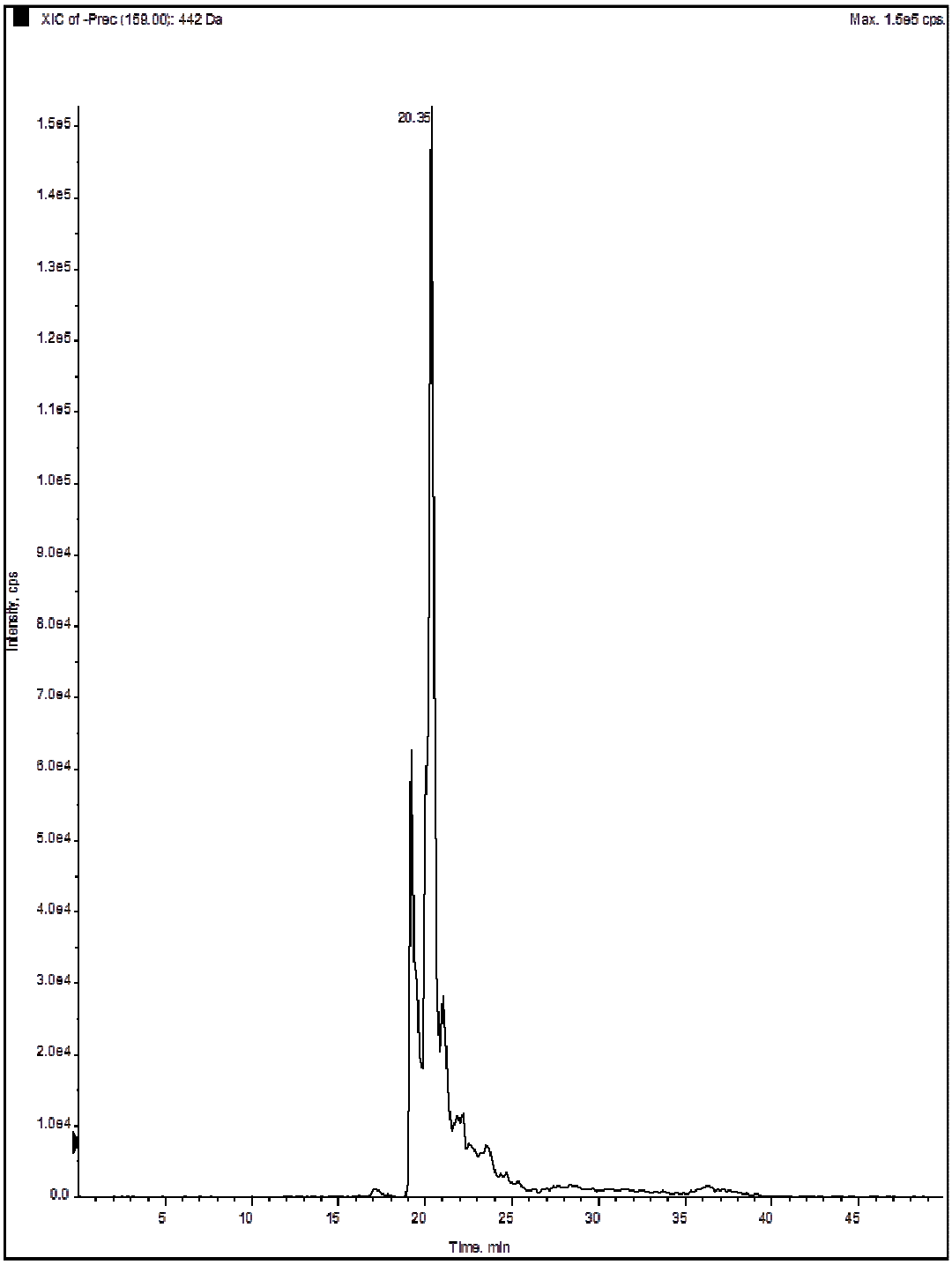
A


XIC of -Prec (159.00): 456 Da
Max. 5494.8 cps.
5495
15.89
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Intensity, cps
5
10
15
20
25
30
35
40
45
Time, min

B



C


XIC of -Prec (159.00): 442 Da
Max. 1.5e5 cps.
20.35
Intensity, cps
1.5e5
1.4e5
1.3e5
1.2e5
1.1e5
1.0e5
9.0e4
8.0e4
7.0e4
6.0e4
5.0e4
4.0e4
3.0e4
2.0e4
1.0e4
0.0
5
10
15
20
25
30
35
40
45
Time, min

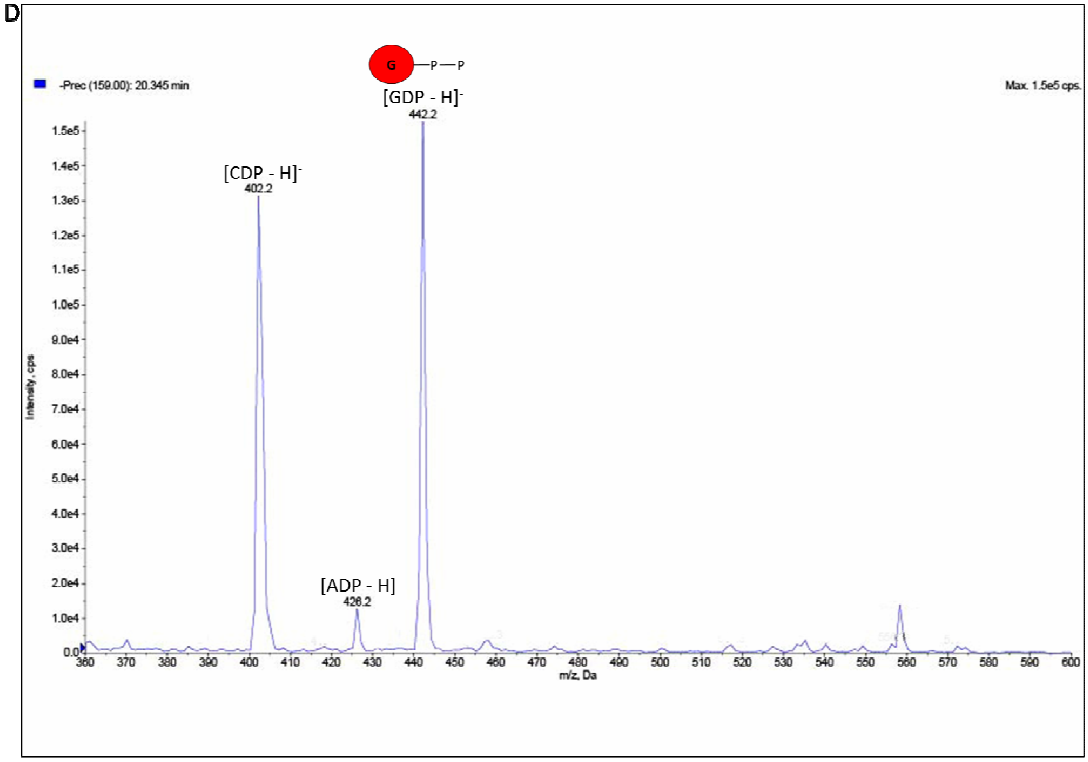


Figure 4.15:*(previous pages)* Precursor ion scan of TFA hydrolysed HEK-293 cell mRNA. Spectrum of the pyrophosphate ion (*m/z* 159) precursor ion scan produced peaks of the predicted pyrophosphate containing precursor ions. (A) Extracted ion chromatogram (XIC) of pyrophosphate precursor ion at *m/z* 456. (B) The peak observed at 15.89 m was identified as the m7GDP $[M-H]^-$ ion. (C) XIC of pyrophosphate precursor ion at *m/z* 442. (D) The peak observed at 20.35 m was identified as the GDP $[M-H]^-$ ion. The CDP $[M-H]^-$ ion at *m/z* 402 and ADP $[M-H]^-$ ion at *m/z* 426 were also detected. Data collected on QTRAP 4000.

RESULTS – CHAPTER III

RESULTS – CHAPTER III

5. Comparative transcriptomics and proteomics of *T. brucei* polysomal mRNA

5.1 Preparation of polysomal RNA

Previous transcriptome analyses using RNA-seq on laboratory cultured *T. brucei* found that gene expression is predominantly controlled post-transcriptionally (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010). Consequently, the significant changes in protein levels observed between life cycle stages cannot be accounted for using transcript abundance alone(Urbaniak et al., 2012, 2013). Given this disparity between transcript and protein ratios, our hypothesis was that mRNA recruitment to ribosomes might contribute to the regulation of gene expression in *T. brucei*. To test this hypothesis, polysome profile analysis was used to monitor the efficiency of translation between bloodstream form and procyclic form *T. brucei*. More than one 80S ribosomal monosome subunit can be translating an mRNA transcript at a time producing so called 'polysomes' (Mašek et al., 2011). The number of ribosomes on an mRNA generally reflects that transcripts translatability under given conditions (Spirin, 1999). Further, a particular mRNA's higher or lower than average association with ribosomes indicates potential involvement of gene-specific regulatory mechanisms (Pradet-Balade et al., 2001). Because translation is a highly dynamic process, cells were treated with the antibiotic cycloheximide to prevent polysome run-off during sample preparation. Cycloheximide binds to the 60S ribosomal subunit and arrests translation elongation by inhibiting release of the deacylatedtRNA from the ribosome E site, thereby stalling the ribosomes on mRNA in a polysomal state (Obrig et al., 1971; Stöcklein and Piepersberg, 1980).

The high protein content of polysomes allows them to be separated throughout a sucrose gradient according to the number of ribosomes attached to the mRNA (Figure 5.1). Free ribosomal subunits and mRNA that is not associated with ribosomes fractionate at the top of the gradient and are known as 'subpolysomes', while polysomes contain transcripts that are associated with at least one ribosome and fractionate in the second half of the gradient. The RNA content and separation of subpolysomes from polysomes across the gradient was monitored by UV absorbance at 254 nm. Following polysome profiling, RNA purified either from individual fractions or pooled polysomal fractions was used for qRT-PCR and RNA-seq analysis (Figure 5.1).

5.2 Polysome profile distributions of developmentally regulated genes

To investigate the relative translation between BSF and PCF *T. brucei*, cell lysates of the two life stages were subjected to polysome gradient fractionation. Equal numbers of cells were treated with cycloheximide, and the clarified lysate was resolved through a sucrose gradient. While both BSF and PCF appeared to be highly translating cells (indicated by there being a greater proportion of RNA in the polysomal fraction of the gradient) PCF cells showed a substantially greater ratio of polysomal to subpolysomal RNA (Figure 5.2). A group of five genes were selected to investigate their translational efficiency on the basis that the difference in their protein expression levels between the two life cycle stages is greater than the observed difference in transcript levels. To understand if ribosome association or segregation plays a role in the regulation of protein expression, the distribution of the transcripts for the genes was analysed across the polysome profile using qRT-PCR.

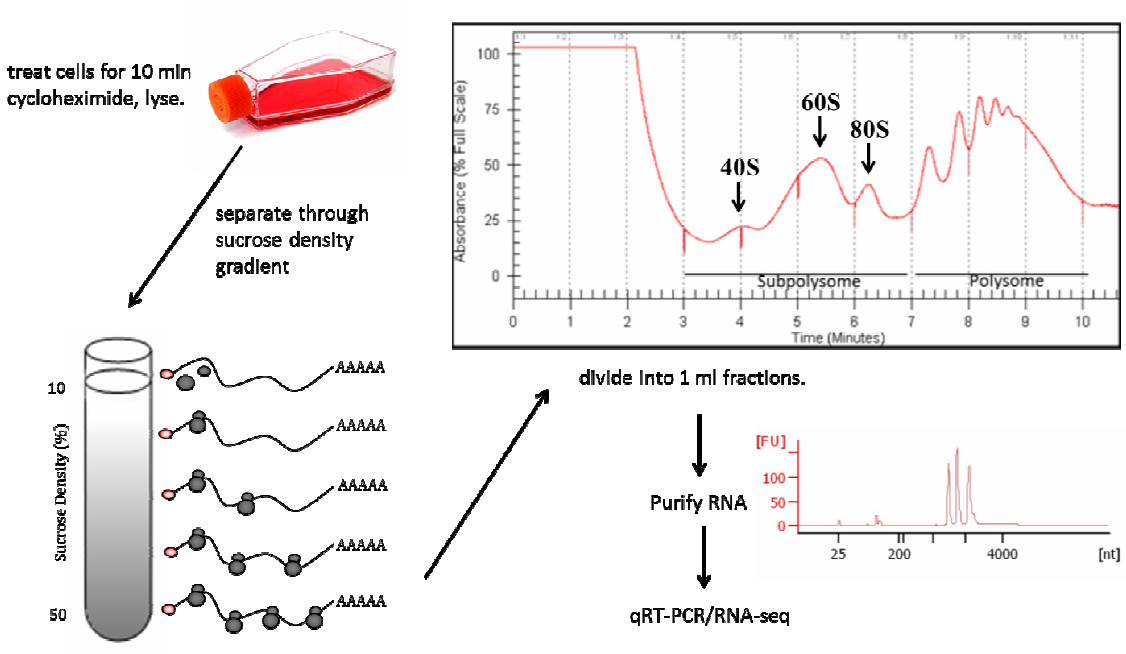


Figure 5.1: Schematic of cell lysate resolution through a sucrose gradient. Representative absorbance trace of a BSF polysome profile during collection of fractions.The first seven fractions contain subpolysomes and the last four fractions contain polysomes.Free monosomes (80S) and ribosomal subunits (40S and 60S) are indicated with arrows. Representative bioanalyserelectropherogram of purified RNA from the gradient depicting the three characteristic rRNA peaks. The baseline of polysome profile was normalized to the absorbance of a blank gradient.

Dolicholphosphate mannose synthase (DPMS) was chosen as a control gene because its function in the formation of the trimannoside core structure of the GPI anchor and the $Man_9GlcNAc_2$-P-P-dolichol precursor of protein N-glycosylation exists in both life cycle stages of *T. brucei*(Mazhari-Tabrizi et al., 1996).As predicted, the distribution of DPMS mRNA was predominantly in the polysomal fraction of the gradient which indicated that was a highly translated transcript in both bloodstream and procyclic form *T. brucei* (Figure 5.3, panel A).RAB18, a GTPase that is important for the processing of surface proteins for vesicle trafficking in the Golgi complex. It has only been observed

in BSF at the protein level, but there is RNA detected in the PCF (Jeffries et al., 2002). Although RAB18 had a greater proportion of its transcripts in the BSF polysome fraction relative to the PCF polysome fraction (Figure 5.3, panel B), the effect was not profound and is unlikely to account for the BSF-specific nature of RAB18 protein expression.

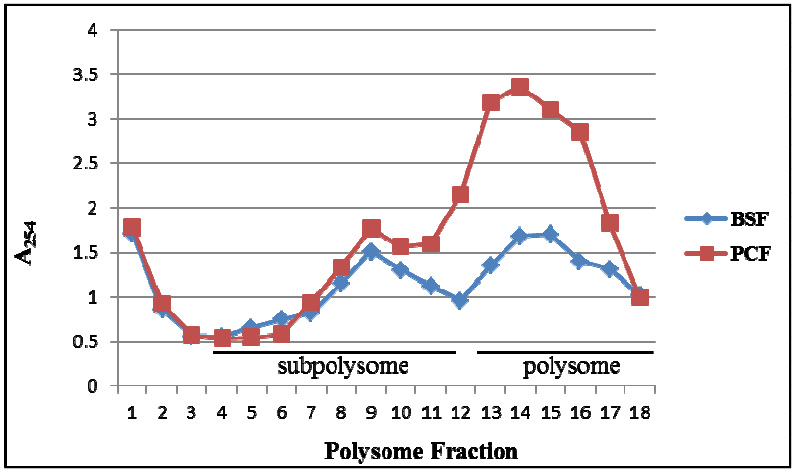


Figure 5.2: Polysome profiles of two different *T. brucei* life cycle stages. Representative absorbance (254 nm) trace of polysome profiles post fractionation comparing BSF and PCF.

Transketolase, a component of the pentose phosphate pathway, is involved in the generation of ribose 5-phosphate for nucleotide biosynthesis and NADPH for defence against oxidative stress. This procyclic form specific protein is undetectable in BSF lysate both by Western blot and in *Tb*TKT enzymatic assays (Stoffel et al., 2011). Although significantly reduced in comparison with PCF, *Tb*TKT transcript was detected in BSF cells by Northern blot suggesting the difference in transcript level may be smaller than protein level (Stoffel et al., 2011). Consistent with those results, there was more *Tb*TKT mRNA observed in the PCF polysomal fractions relative to BSF while the

opposite was seen in the subpolysomal fraction (Figure 5.3, panel C). Aconitase is a protein responsible for the conversion of citrate to isocitrate in the Kreb's cycle. It has been shown to be a developmentally regulated gene that is only expressed in the stumpy and procyclic forms. At the RNA level, there is little difference between the life cycle stages suggesting the mechanism of regulation is translational or post-translational (Saas et al., 2000). When analysed by polysome profile, although the procyclic form distribution was ambiguous, there appeared to be more polysomal mRNA in bloodstream form than procyclic form (Figure 5.3, panel D). Another procyclic form specific Kreb's cycle protein, mitochondrial malate dehydrogenase (mMDH), functions in aspartate metabolism by reducing oxaloacetate to malate and enabling the transfer of reducing equivalents (such as NADH) from the cytosol to the mitochondrion (Anderson et al., 1998; Aranda et al., 2006). Similar to the profile obtained for *Tb*TKT, there was more *Tb*mMDH mRNA observed in the PCF polysomal fractions relative to BSF polysomal fractions of the polysome profile (Figure 5.3, panel E). A cytoskeletal calpain-related protein CAP5.5 that is essential for correct procyclic form morphogenesis was not detectable in bloodstream form cells while expression of *CAP5.5* gave evidence for only 5-fold difference in mRNA (Hertz-Fowler et al., 2001; Olego-Fernandez et al., 2009). Polysome profile analysis also showed an increase in polysome associated transcript in PCF compared to BSF (Figure 5.3, panel F). In order to further analyse the varying distribution of these six gene transcripts across a polysome profile, RNA purified from BSF and PCF pooled subpolysomal and polysomal fractions was transcribed into cDNA and analysed by qRT-PCR using gene-specific primers. *TbRAB18, TbTKT, TbmMDH,* and *TbTCAP5.5* all confirmed their respective enrichment in polysome-derived material although (except for *TbTKT*) the increase was not substantial (Figure 5.4). The control gene, *TbDPMS*, showed a slight enrichment in PCF polysomes, while *TbAconitase*showed a slight enrichment in BSF

polysomes, contradicting the results in Figure 5.3. The data for this small number of transcripts appeared mostly promising, however it would be difficult to establish trends for larger subsets of genes based on these methods and therefore a genome-wide approach using RNA-seq was pursued.

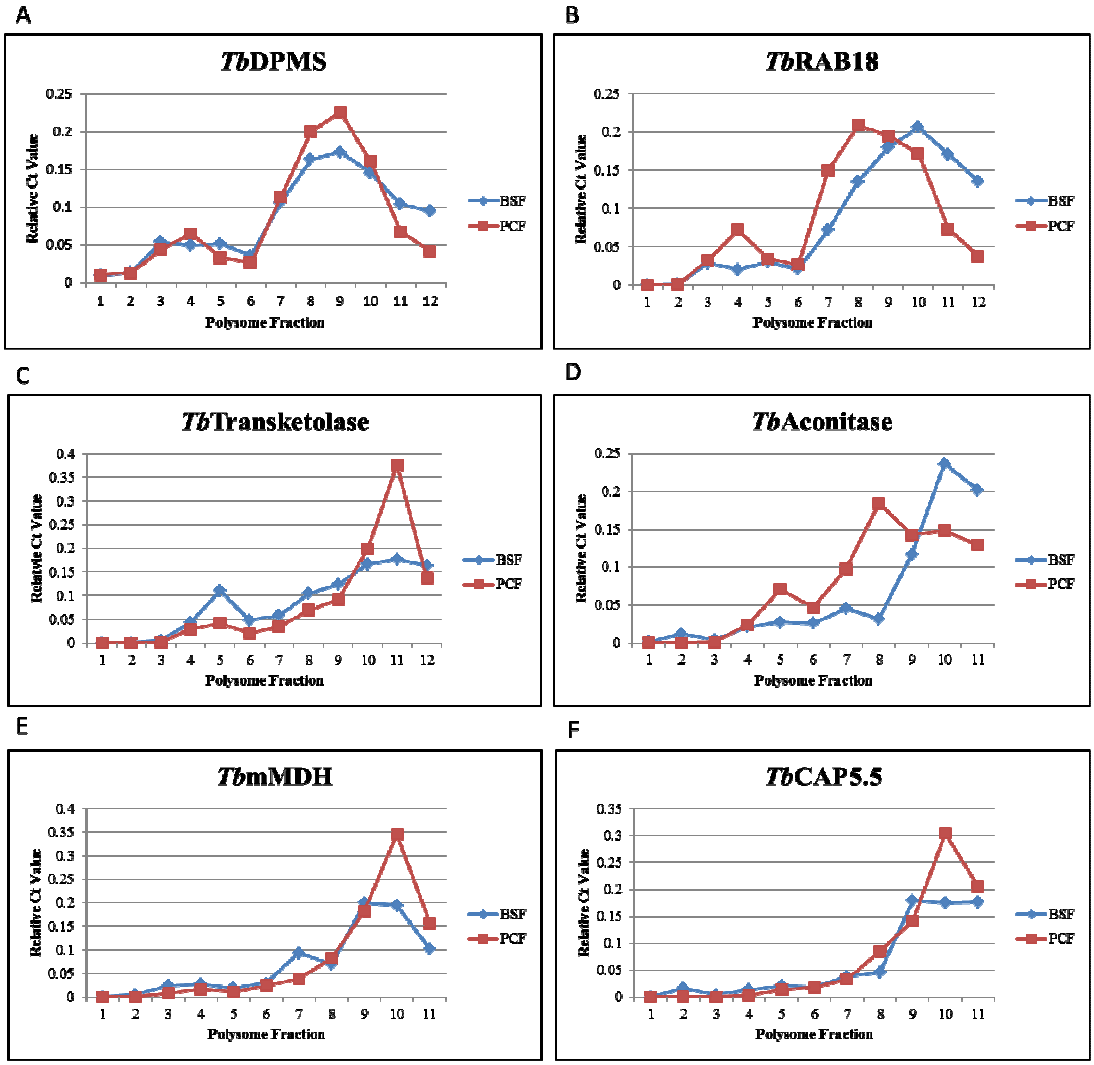


Figure 5.3: Transcript distribution across polysome profiles of bloodstream and procyclic form *T. brucei*. Total RNA was purified from each fraction of two polysome profiles comparing BSF and PCF. qRT-PCR was performed on six transcripts using gene-specific primers: (A) dolicholphosphate mannose synthase (DPMS), (B) RAB18, (C) transketolase, (D) aconitase, (E) mitochondrial malate dehydrogenase (mMDH), and (F) CAP5.5 to analyse their distribution across the sucrose gradient. Genes with the greatest amount of transcript in the later fractions indicates polysome association.

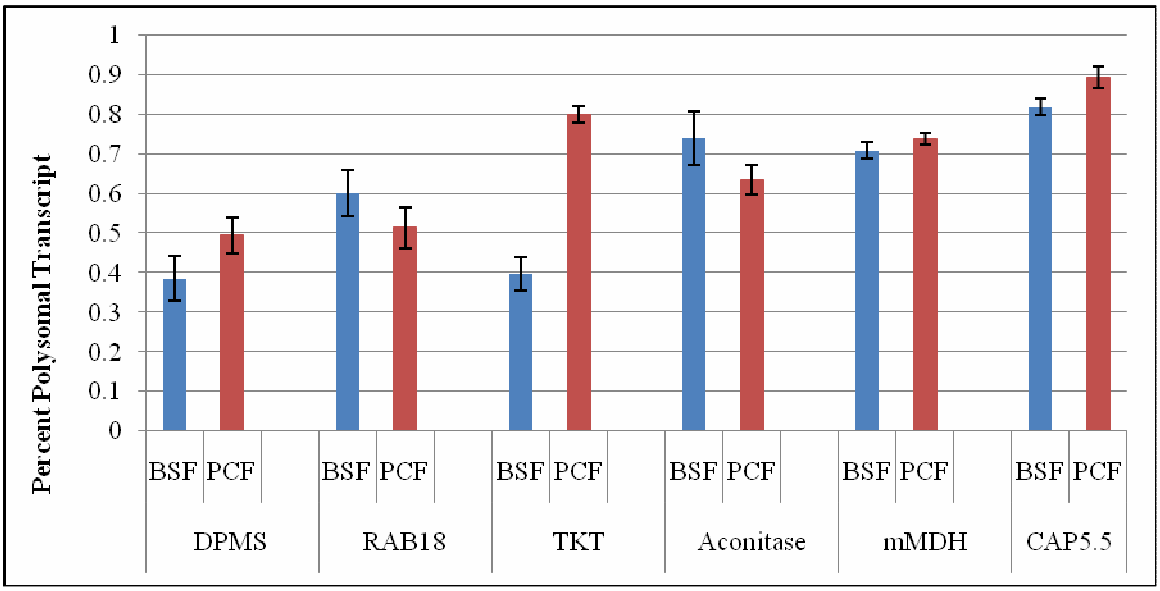


Figure 5.4: Polysomal enrichment ofmRNAs.Total RNA from pooled subpolysomal and polysomal fractions was purified and analysed by qRT-PCR for six candidate genes: dolicholphosphate mannose synthase (DPMS), RAB18, transketolase, aconitase, mitochondrial malate dehydrogenase (mMDH), and CAP5.5 to analyse their relative abundance in the polysomal fraction. Three biological replicates were analysed; the mean ± standard deviations are shown.

5.3 cDNA library preparation for Illumina sequencing

To prepare samples for RNA-seq, cDNA libraries were generated from both total mRNA and polysome-associated mRNA transcripts. Due to the high proportion of rRNA in a cell (>90%), it was depleted from the sample using oligo-dT beads which enrich for polyadenylated mRNA. The removal of rRNA is essential to maximize coverage of low abundance transcripts from the number of reads possible per machine run. Three possible mRNA enrichment strategies were tried: Illuminaoligo-dT beads, Dynaoligo-dT beads, and treatment with Terminator 5´-Phosphate-Dependent Exonuclease. The Illuminaoligo-dT bead method produced the most consistent results and was the method of choice for subsequent cDNA library preparation (data not shown).

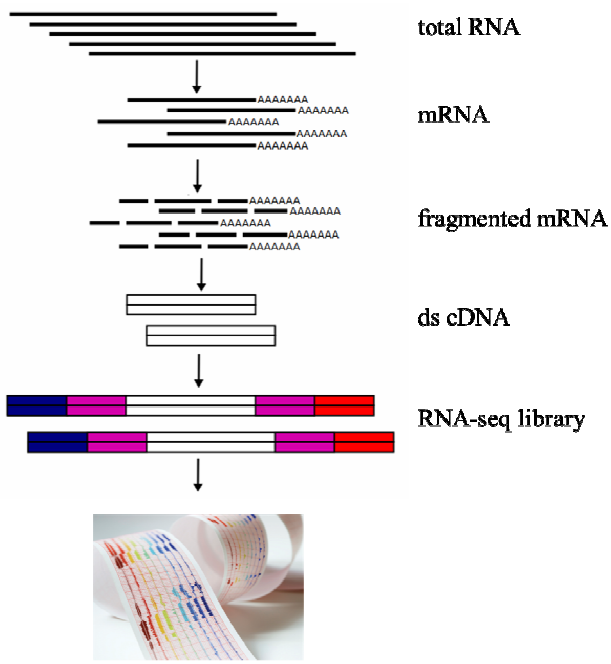


Figure 5.5: cDNA library preparation method. Schematic of the process of cDNA library generation: total RNA extracted from *T. brucei* cells is enriched for mRNA, fragmented and transcribed into DNA with a reverse transcriptase enzyme. Bar-coded adaptors are ligated to the double stranded cDNA to produce the final library to be sequenced.

Following purification, the mRNA was sheared into smaller fragments for optimal sequencing coverage of smaller transcripts and to circumvent the potential bias of reverse transcriptase. Ultrasonic fragmentation conditions were optimized to produce ~200 nucleotide sections to maximize coverage of the sequencer's paired-end reads (data not shown). The fragmented mRNA was then reverse transcribed into cDNA. Double stranded cDNA was then end repaired, and a terminal dA nucleotide was added to prime for the adaptor ligation, resulting in the final cDNA library to be sequenced (Figure 5.5). It is important to note that unlike the established cDNA library protocol, our procedure enabled the library to be completed without PCR amplification, therefore eliminating the sample bias associated with variable amplification. Throughout the cDNA library preparation procedure, samples were monitored by Agilent Bioanalyser to ensure RNA integrity and cDNA quantity (Figure 5.6).

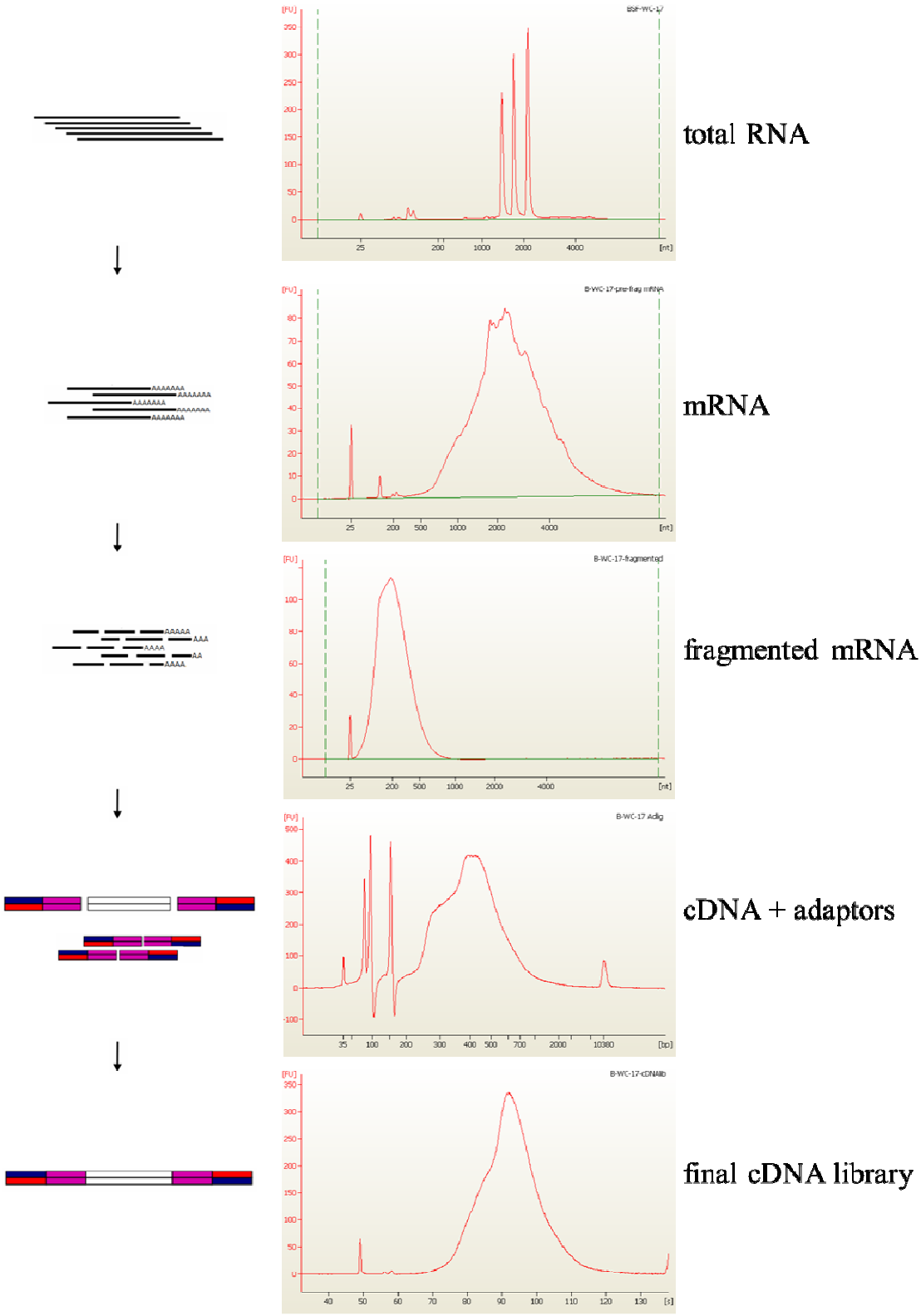
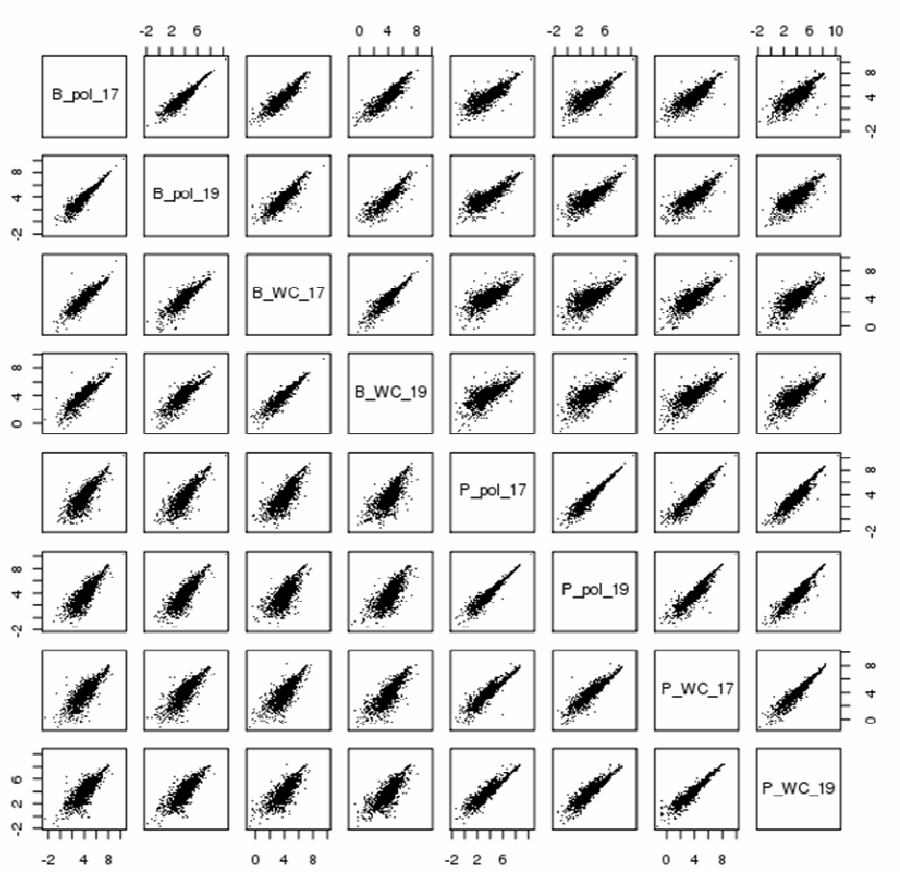


Figure 5.6: Representative Bioanalyzerelectropherograms of cDNA libraries prepared for Illumina sequencing from *T. brucei* total and polysome associated RNA.RNA size in nucleotides (nt) is depicted on the x-axis and fluorescence units (FU) on the y-axis. The peak visible at 25 nt corresponds to the RNA/DNA sample dye.

5.4 RNA-seq of whole-cell mRNA and polysome associated mRNA in bloodstream and procyclic form *T. brucei*

In collaboration with the Wellcome Trust Sanger Institute, eight libraries (2 biological replicates of BSF and PCF polysomal and total (whole-cell) derived mRNA) were constructed for a preliminary RNA-seq experiment to begin to investigate the hypothesis that mRNA selection by ribosomes might contribute to the regulation of gene expression.Using next-generation sequencing (IlluminaMiSeq), we obtained 16.6 million high-quality reads (8.3 million read-pairs) for the first technical replicate and 18.2 million high-quality reads (9.1 million read pairs) for the second technical replicate. A high correlation for the read counts of biological and technical replicates (average Pearson correlation coefficient R = 0.967) confirmed the reproducibility of the methodology (Figure 5.7, panels A and B). Scatter plots of the correlations between the log of Reads Per Kilobase of exon model per Million mapped reads (RPKM) values for the two biological replicates of the two conditions examined in this experiment (whole cell mRNA (WC) and polysome associated mRNA (pol)) also depicted the differences between bloodstream and procyclic form samples as well as between total mRNA and polysomal mRNA samples. For low RPKM values, (lower left corner of graphs) the slight pear-shape was to be expected due to the increase in errors for low read signals (Figure 5.7, panel A). The differences of the mRNA populations in the samples analysed were appreciated by the numbers of significantly differentially expressed transcripts (< p=0.00001) (Figure 5.7, panel C). The biggest difference in gene expression was between bloodstream and procyclic form total mRNA (805 transcripts). Bloodstream and procyclic polysomal mRNA also revealed a similar number of translational changes (755 transcripts). Within each life cycle stage, fewer genes were seen with changes between total mRNA and polysomal mRNA (BSF: 241 transcripts, PCF: 144 transcripts).

A



B

RPKM pairwise Pearson correlations

	B_pol_17	B_pol_19	B_WC_17	B_WC_19	P_pol_17	P_pol_19	P_WC_17	P_WC_19
B_pol_17	1.0000	0.9653	0.9468	0.9500	0.9147	0.9145	0.9165	0.9121
B_pol_19		1.0000	0.9498	0.9516	0.9198	0.9199	0.9264	0.9186
B_WC_17			1.0000	0.9678	0.8847	0.8854	0.9217	0.9117
B_WC_19				1.0000	0.8841	0.8863	0.9254	0.9151
P_pol_17					1.0000	0.9688	0.9496	0.9454
P_pol_19						1.0000	0.9492	0.9477
P_WC_17							1.0000	0.9679
P_WC_19								1.0000

C

	BSFtotal	BSFpolys	Ptotal	Ppolys
BSFtotal	x	241	805	
BSFpolys		x		755
Ptotal			x	144
Ppolys				x

adjusted p=0.00001

Legend:
B = bloodstream form
P = procyclic form
wc = whole cell (total) mRNA
pol = polysomal mRNA
17 and 19 = indicate the two biological replicates

Figure 5.7: RNA-seq analysis. (A) Matrix of pairwise scatter plots of the correlation between log Reads Per Kilobase of exon model per Million mapped reads(RPKM) values between two biological replicates of two conditions (whole cell (WC) mRNA and polysomal (pol) mRNA). (B) Table of pairwise Pearson correlation values for comparisons in panel A. (C) Table of significantly differentially expressed (< p=0.00001) transcripts between two life cycle stages (BSF, PCF) and two conditions (whole cell mRNA, polysomal mRNA). Graphs and tables generated by Bernardo Foth, Sanger Institute.

Encouraged by the reproducibility of the replicates and potentially interesting differences between mRNA populations, a larger RNA-seq experiment was pursued.

5.5 Comparative transcriptomics by RNA-seq of total, subpolysomal, and polyribosomal poly (A+) RNA purified from bloodstream form and procyclic form *T. brucei*

Following the preliminary experiment described above, three biological and three technical replicates of total, subpolysomal, and polysomal mRNA were sequenced using an Illumina High-Seq 2000 sequencer. The three technical replicates showed nearly perfect reproducibility and were therefore merged for each biological replicate. The reproducibility of the biological replicates was also deemed reliable. Four different analysis pipelines confirmed that there were hundreds of genes that were significantly differentially expressed and the overlap between the sets of genes generated by these different analyses was well over 90%. The Pearson correlations were calculated between the ~8400 data points in pairwise comparison to assess concordance within the sample replicates, yielding an average 99% correlation (Table 5.2). There were between 1182 and 4501 significantly differentially expressed/polysome-associated genes in all possible pairwise sample comparisons (Table 5.1). The smallest number of differentially expressed genes (1182) was between the BSF and PCF subpolysome samples; while the largest number of differentially expressed genes (4501) was between the BSF whole cell and BSF subpolysome samples.

Such large numbers of genes and the complexity of biological networks make clustering a useful precursory technique for analysing gene expression data. Principle component analysis (PCA) was used to visualize the interrelationships between the different RNA-seq samples (Figure 5.8). The PCA confirmed the very high reproducibility of the technical replicates and showed the concern over sample procyclic form subpolysomal

replicate #3 (PCFsub-3) because it showed the most variation in its biological replicates. Further enquiry was necessary to understand the source of the outliers (Figure 5.8). In PCFsub-3, there were 375 genes for which the normalized read count deviated by more than 1.5-fold from the average of the other 2 replicates (PCFsub-1 and PCFsub-2). Of these genes, 95 were contaminant RNAs (76x snoRNAs, 19x ribosomal RNAs) and 151 of the genes were of unknown function (annotated as hypothetical in tritryp genome database). Slight variation in the efficiency with which RNAs other than mRNAs are excluded during cDNA library preparation would easily show up in such comparisons between replicate samples. The PCFsub-3 sample was compared to a sample where the biological replicates appeared to be very close together (PCFpol) and altogether, the number of outlier genes in PCFsub-3 was only moderately increased compared to other biological samples. Therefore, PCFsub-3 was included in all analyses with adjusted p-value$<1^{-10}$stringency.

	read mapping:	Tophat
	differential analysis	edgeR
	min. adj. p-value:	1e-10
Pairwise comparison		
	Bpol_vs_Bsub	3238
	Bpol_vs_Ppol	2769
	Bsub_vs_Psub	1182
	Bwc_vs_Bpol	2195
	Bwc_vs_Bsub	4501
	Bwc_vs_Pwc	3408
	Ppol_vs_Psub	3843
	Pwc_vs_Ppol	2354
	Pwc_vs_Psub	3044

Table 5.1: Numbers of significantly differentially expressed/polysome-associated genes. Pairwise comparison of the six conditions studied (total, subpolysomal and polysomal mRNA from BSF and PCF *T. brucei*). Reads were mapped to the reference genome using SMALT software and differential expression analysis was completed using DESeq software. Analysis performed by Bernardo Foth, Sanger Institute.

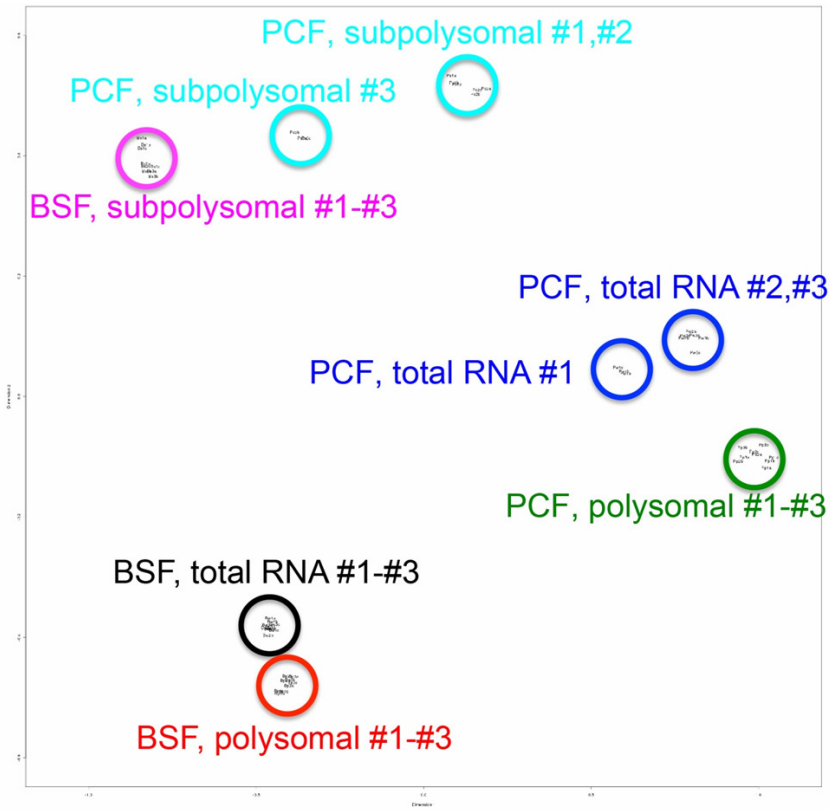


Figure 5.8: Principle component analysis (PCA) of RNA-seq data set. PCA of the RPKM values between three biological and three technical replicates of six conditions: total, subpolysomal and polysomal mRNA from BSF and PCF *T. brucei*. Clusters of the data from each of the six conditions are outlined with circles. The axes of the graph are arbitrary units of the principle components used to perform the analysis. Graph generated by Bernardo Foth, Sanger Institute.

	biological replicate		
mRNA sample	1 vs 2	2 vs 3	1 vs 3
Pwc	99.6	99.7	99.4
Ppol	99.7	99.8	99.8
Psub	97.1	98.7	97.3
Bwc	99.8	99.8	99.8
Bpol	98.8	99.7	99
Bsub	98.9	97.9	97.9

average: 99.04%

Table 5.2: Pearson correlations of the RNA-seq biological replicates. Pairwise comparisons between the ~8400 data points were used to calculate the corresponding Pearson correlations expressed as a percentage. Correlations generated by Bernardo Foth, Sanger Institute.

5.6 Comparison of RNA-seq total and polysome associated transcript levels between bloodstream and procyclic form T. brucei

The data set was then used to address the question as to whether translational efficiency might be regulated in a life cycle-specific manner. To this end, the BSF:PCFlog-fold changes of normalized read counts for total mRNA were compared to those for polysome associated mRNA.This revealed a strong positive correlation (Pearson correlation coefficient = 0.95) (Figure 5.9) suggesting that the majority of transcripts are similarly distributed between the total mRNA and polysomal mRNA fractions. Examining the differential expression of total mRNA between the life stages, we observed the greatest up-regulation of procyclin transcripts in procyclic form and VSGs and glucose transporter 1B (THT1) transcripts in bloodstream form (Figure 5.10, panel A). Similarly, the greatest differences in polysome associated transcripts were observed in procyclins for procyclic form and VSGs and THT1 for bloodstream form (Figure 5.10, panels C). Analysis of the translational efficiency of transcripts within individual life stages showed that the most polysome-associated transcripts in bloodstream form were for the proteolysis pathway and translation machinery (ribosomal mRNA) and the most abundant transcripts in the total mRNA fraction were in fact contaminating rRNAs (Figure 5.11, panel A). Unexpectedly, in procyclic form, no transcripts showed enrichment in the polysomal fraction relative to the total mRNA fraction and the most abundant transcripts in total mRNA were also contaminating rRNA (Figure 5.11, B). These results suggested translational efficiency does not significantly contribute to the regulation of gene expression in *T. brucei*.

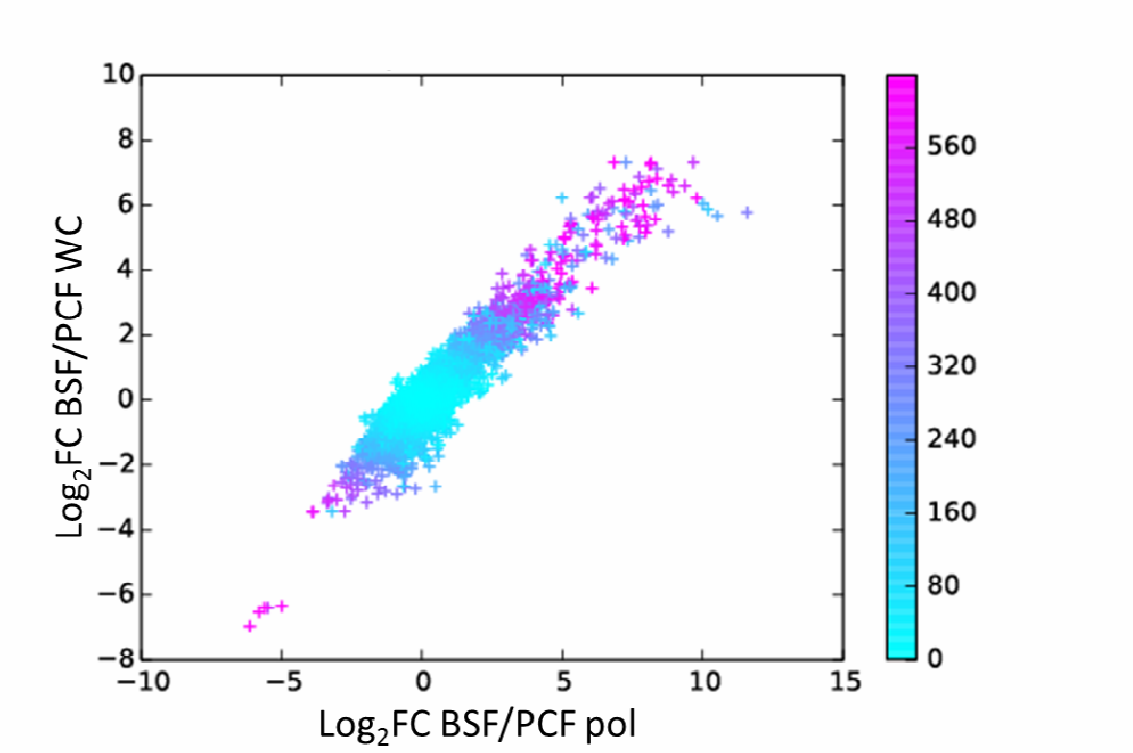
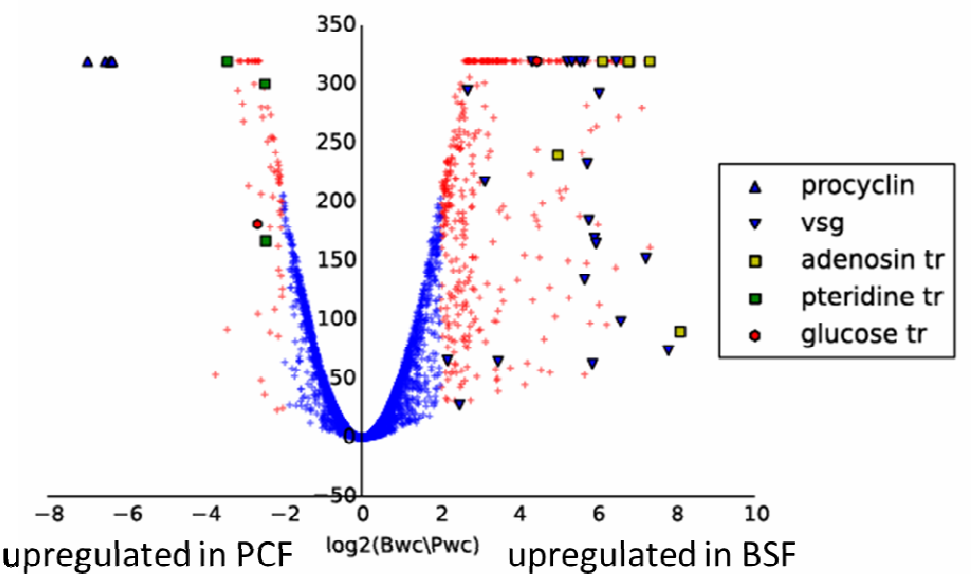


Figure 5.9: Scatter plot showing the comparison of normalized log fold changes. The log fold changes between BSF and PCF polysome associated transcripts (pol) was compared to the log fold changes of whole cell derived total mRNA (WC). Pearson correlation=0.95 Thecolored gradient denotes the $\log_{10}$ of the sum of the p-values for each correlation pair (i.e. pink correlation points represent the highest significance). Graph generated by Michele Tinti, MAJF lab.

A



B

upregulated in BSF

GO term Group	Frequency	Description
GO:0045121	13/14 (92.86%)	membrane raft
GO:0009986	13/26 (50.00%)	cell surface
GO:0003674	54/758 (7.12%)	molecular_function
GO:0007155	7/12 (58.33%)	cell adhesion
GO:0004016	12/58 (20.69%)	adenylate cyclase activity
GO:0008150	53/834 (6.35%)	biological_process
GO:0009190	12/68 (17.65%)	cyclic nucleotide biosynthetic process
GO:0016849	12/68 (17.65%)	phosphorus-oxygen lyase activity
GO:0035556	12/82 (14.63%)	intracellular signal transduction
GO:0016021	31/443 (7.00%)	integral to membrane
GO:0004222	7/30 (23.33%)	metalloendopeptidase activity
GO:0005886	6/23 (26.09%)	plasma membrane

upregulated in PCF

GO term Group	Frequency	Description
GO:0004308	3/9 (33.33%)	exo-alpha-sialidase activity
GO:0009405	3/10 (30.00%)	pathogenesis
GO:0020012	3/10 (30.00%)	evasion or tolerance of host immune response
GO:mik00003	3/16 (18.75%)	procyclin

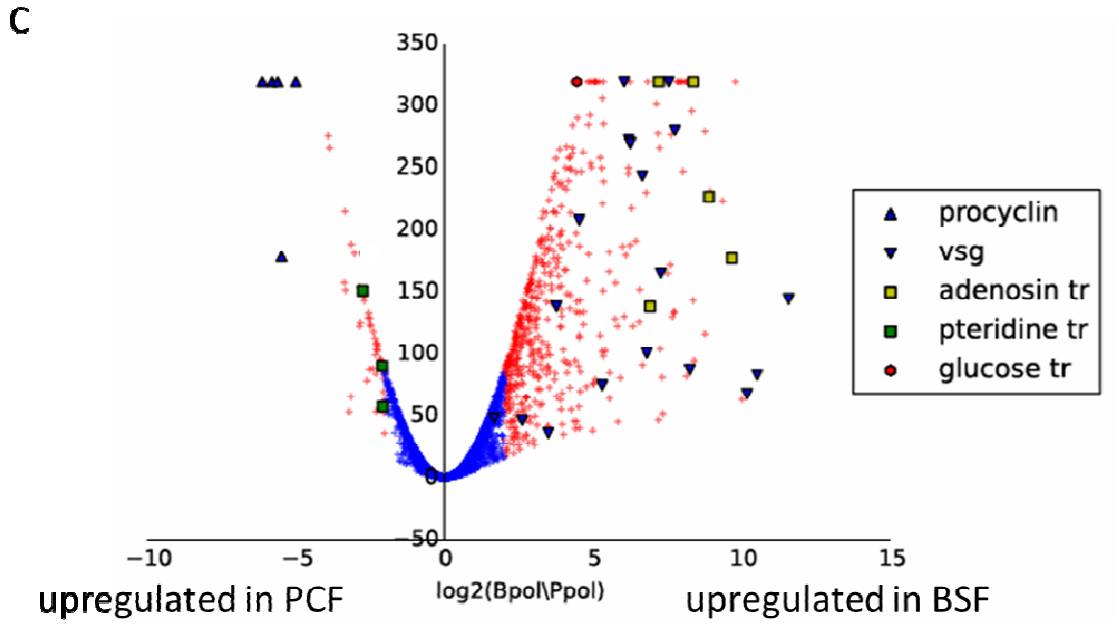
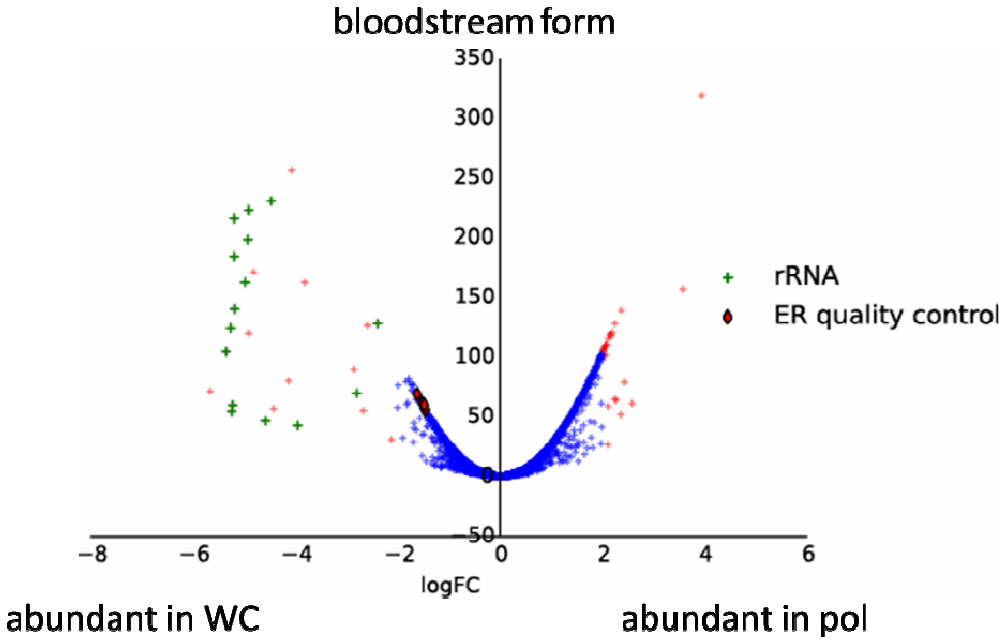


Figure 5.10:Volcano plots of (A) whole cell derived (WC) mRNA and (C) polysomal (pol) mRNA in bloodstream versus procyclic form *T. brucei*. The horizontal axis is the $\log_2$ fold change between the life cycle stages. The $-\log_{10}$ (P-value) is on the vertical axis. Each transcript is represented by one point on the graph. Points with positive values on the horizontal axis are up-regulated in bloodstream form and points with negative values are up-regulated in procyclic form. Red points highlight genes with > 4 fold differences in transcript abundance ($p<0.05$). (B) Gene ontology (GO) term analysis of pathways whose components exhibited elevated expression in BSF whole cell mRNA. Graphs generated by Michele Tinti, MAJF lab.

A


bloodstream form
350
300
250
200
150
100
50
0
−50
−8
−6
−4
−2
0
2
4
6
logFC
rRNA
ER quality control
abundant in WC
abundant in pol

B

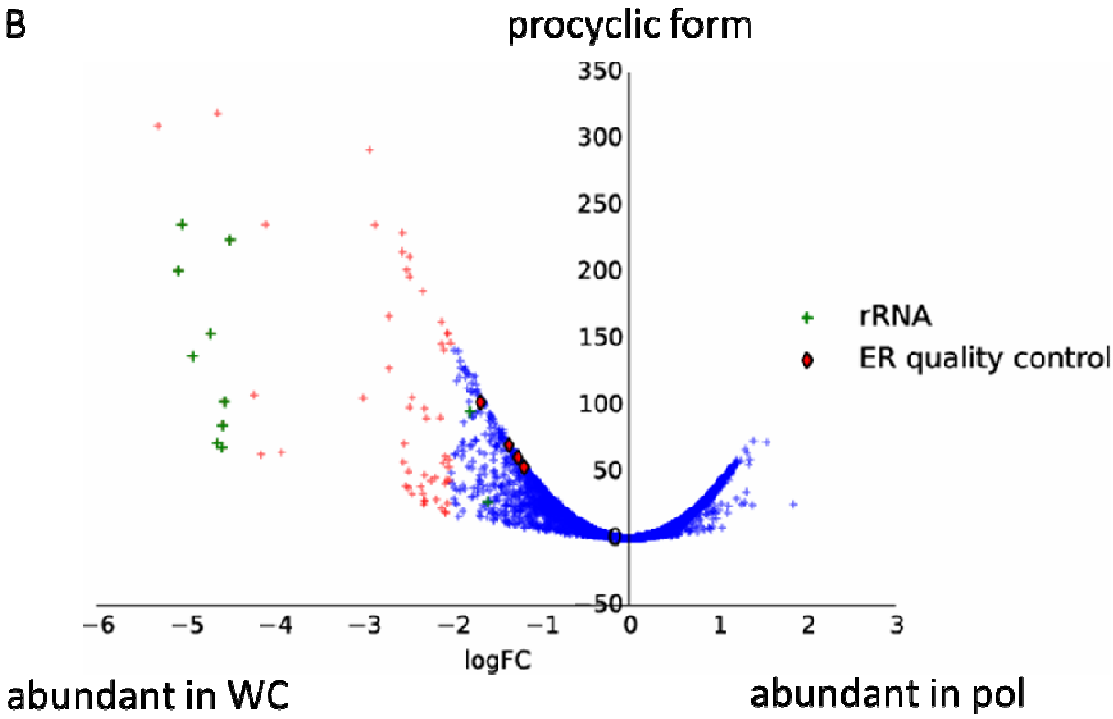

procyclic form
350
300
250
200
150
100
50
0
−50
−6
−5
−4
−3
−2
−1
0
1
2
3
logFC
rRNA
ER quality control
abundant in WC
abundant in pol

Figure 5.11 *(previous page)*: Volcano plots of (A) bloodstream form and (B) procyclic form *T. brucei* whole cell derived (WC) mRNA versus polysomal (pol) mRNA. The horizontal axis is the $\log_2$ fold change between WC and pol transcripts. The $-\log_{10}$ (P-value) is on the vertical axis. The absolute expression value (RPKM) for each transcript is represented by one point on the graph.rRNA contaminants are designated in green and genes involved in the ER quality control pathway are designated as red diamonds. Points with positive values on the horizontal axis are more abundant in polysome associated mRNA and points with negative values are more abundant in WC mRNA. Graph generated by Michele Tinti, MAJF lab.

5.7 Analysis of RNA-seqdata for total, subpolysomal, and polyribosomal poly (A+) RNA from bloodstream form and procyclic form T. brucei

The phosphoglycerate kinase (PGK) genes are a good example of post-transcriptionally controlled, developmentally regulated genes. Although they are in a tandem array and transcribed in a single polycistron, PGKA is expressed at a low level in both life stages, PGKB is only expressed in the procyclic form and PGKC is only expressed in bloodstream form *T. brucei* (Figure 5.12, panel A) (Blattner and Clayton, 1995; Colasante et al., 2007; Kapotas and Bellofatto, 1993). Consistent with those findings, PGKC transcripts were found in the whole cell (WC) and polysomal RNA fractions of bloodstream form cells; PGKB transcripts were found in the WC and polysomal RNA fractions of procyclic form cells, and PGKA transcripts were found in both bloodstream and procyclic form fractions, but at lower expression levels (Figure 5.12, panel B).

Several mRNA transcripts were observed to be under-represented on polysomes: i.e. the level of transcript in the total mRNA sample was greater than the level of transcript in the polysomal fraction. This suggested that for certain genes, there may be a portion of transcripts that exist in the cells but are not translated. A number of proteins

(calreticulin, BiP (Grp78), glucosidase II, UGGT, and PDI (ERp57)) involved in the endoplasmic reticulum (ER) quality control pathway displayed the "polysome under-representation" phenotype (Figure 5.13). In most eukaryotes, the ER quality control pathway initiates followingdeglucosylation of newly added $Glc_3Man_9GlcNAc_2$ structures by alpha-glucosidasesI and II. In the case of trypanosomes which only transfer $Man_9GlcNAc_2$ and $Man_5GlcNAc_2$, this inital trimming of α-Glc residues does not occur. Correctly folded glycoproteins proceed through the ER and to the Golgi apparatus (Trombetta et al., 1996). Misfolded glycoproteins, however, are glucosylated by the UDP-Glc:glycoproteinglucosyltransferase (UGGT). Calreticulin is a chaperone with a carbohydrate-binding domain that recognizes the UGGT monoglucosylatedglycans on the misfolded glycoproteins and forms a network with other ER chaperones including, BiP and PDI, that function to re-fold the glycoproteins correctly (Tatu, 1997; Vassilakos et al., 1998). As UGGT exclusively glucosylates misfolded glycoproteins, its function serves as a sensor for glycoprotein conformation (Sousa et al., 1992). Upon reaching theirtrue tertiary structure, the glycoproteins are deglucosylated by alpha-glucosidase II and are then permitted to continue to the Golgi apparatus. However, when processing by the folding machinery in the ER is insufficient to produce the native conformation of the glycoproteins, they are ultimately degraded by ER-associated degradation (ERAD) (Figure 5.14) (Tsai et al., 2002).

We hypothesized that cells may use polysome recruitment of transcripts as a reserve mechanism for situations requiring rapid increase in protein expression, such as ER stress. To test this, cells were exposed to two ER stress inducing agents for an hour: dithiothreitol (DTT) a chemical that reduces disulphide bonds and tunicamycin, an inhibitor of N-linked glycosylation which is required for optimal protein folding (Back et al., 2005; Casero et al., 1982).

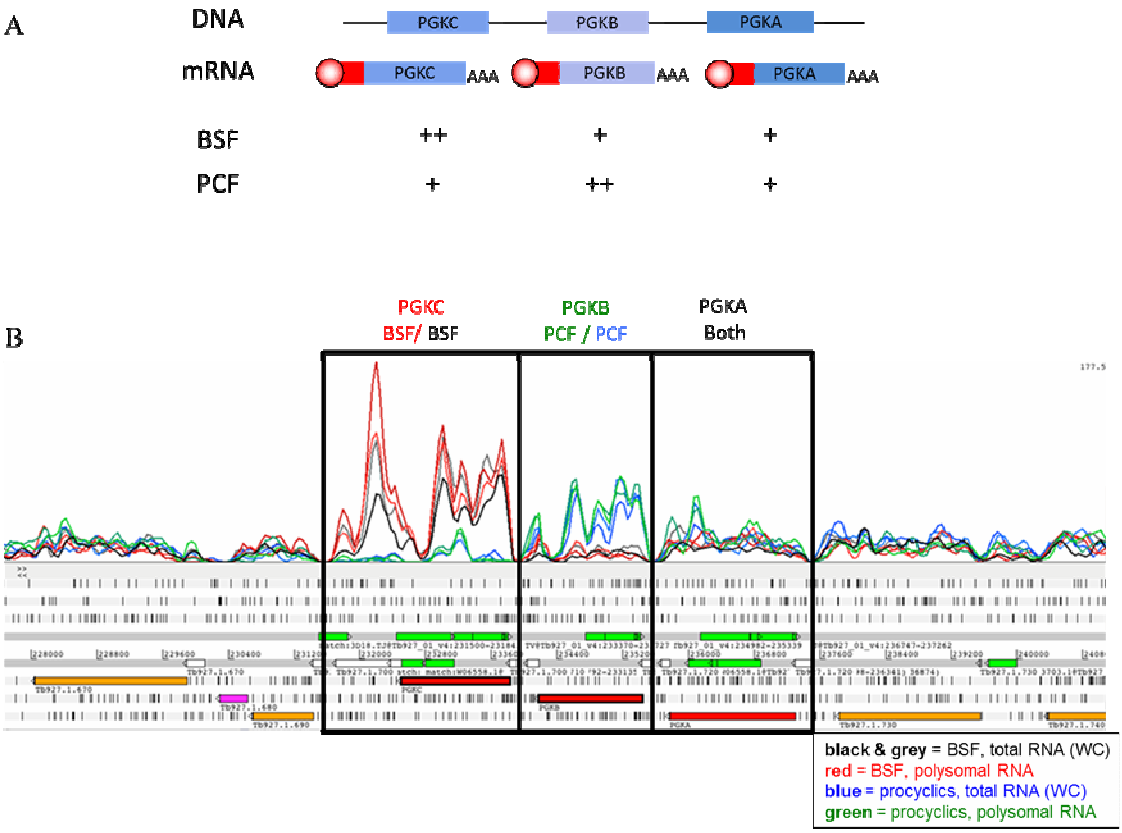


Figure 5.12: The PGK gene locus in *T. brucei*. (A) Genomic arrangement and schematic of post-transcriptional processing and protein abundance. (B) RNA-seq coverage plots of PGKA/B/C mRNA. The number of sequence reads (expressed as the RPKM value) is shown along the length of each gene.

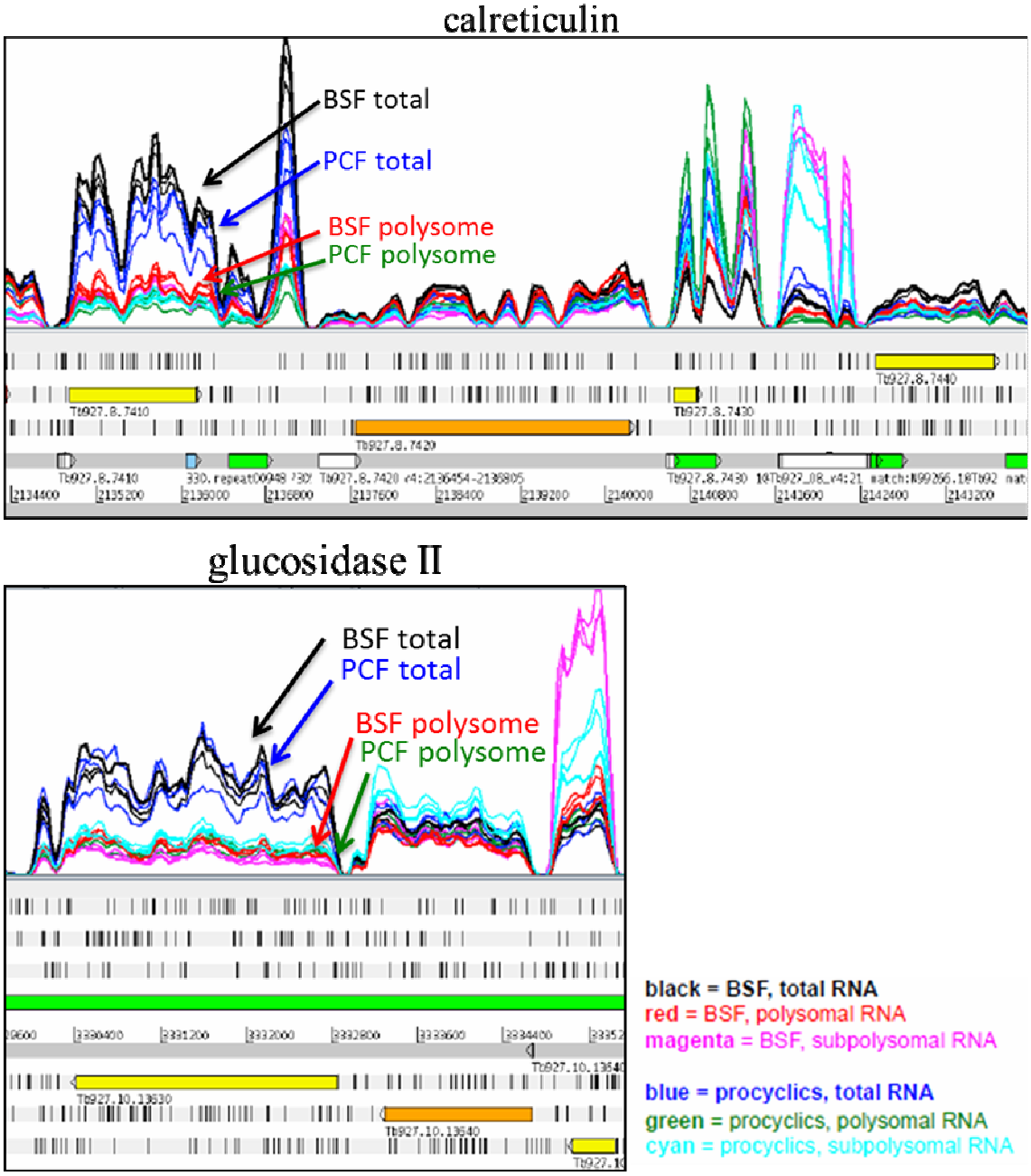


Figure 5.13: RNA-seq coverage plots of calreticulin and glucosidase II mRNA. The number of sequence reads (expressed as the RPKM value) is shown along the length of each gene.

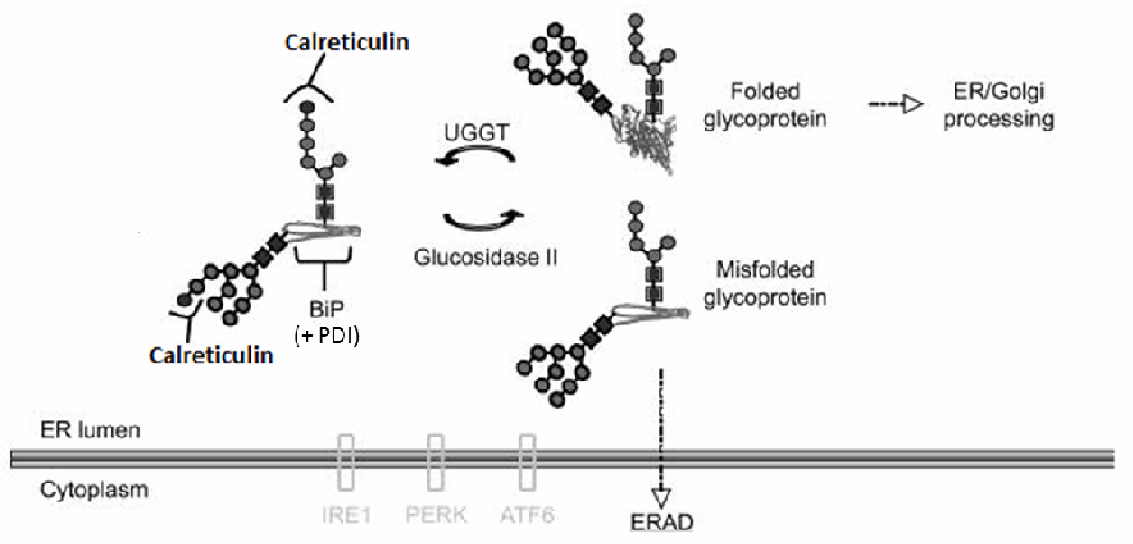


Figure 5.14: The factors involved in the ER quality control pathway in bloodstream form *T. brucei*. Proteins in grey are components of the unfolded protein response (UPR) found in other eukaryotes. BiP: immunoglobulin-heavy-chain-binding protein (Grp78), PDI: protein disulphide isomerase (ERp57), UGGT: UDP-Glc:glycoproteinglucosyltransferase, ERAD: ER-associated degradation. Image adopted from (Izquierdo et al., 2009).

It was found that tunicamycin treatment had undetectable effects on the overall polysome distribution of both bloodstream and procyclic form cells. Both the dimethyl sulphoxide (DMSO) treated control and tunicamycin treated cells showed a typical polysome profile (Figure 5.15). However, treatment with DTT resulted in substantial reduction in global translation efficiency seen in the increased subpolysomal peak and consequently lowered polysomal peaks in the polysome profiles of both bloodstream and procyclic form cells (Figure 5.16).

A subset of mRNAs that are important for the response to the ER stresses were analysed for their potential to redistribute into the polysome peak to enhance their rate of

translation without necessitating additional transcription. Based on the observation of polysome under-representation by RNA-seq, five ER quality control proteins were analysed by qRT-PCR for the translocation of transcripts into the polysomal fraction of RNA following DTT and tunicamycin treatment. Total RNA derived from treated or untreated BSF and PCF *T brucei* was transcribed into cDNA and analysed by qRT-PCR using gene-specific primers. *TbBiP, TbglucosidaseII, TbUGGT, Tbcalreticulin,* and *TbPDI*all showed enrichment in polysome-derived material following tunicamycin treatment even though (except for *TbPDI*) the increase was not substantial (Figure 5.17). Following DTT treatment, the increase in the enrichment of polysomal transcript was even less pronounced, with the exception of *TbPDI* in BSF cells which showed a substantial shift (Figure 5.18).

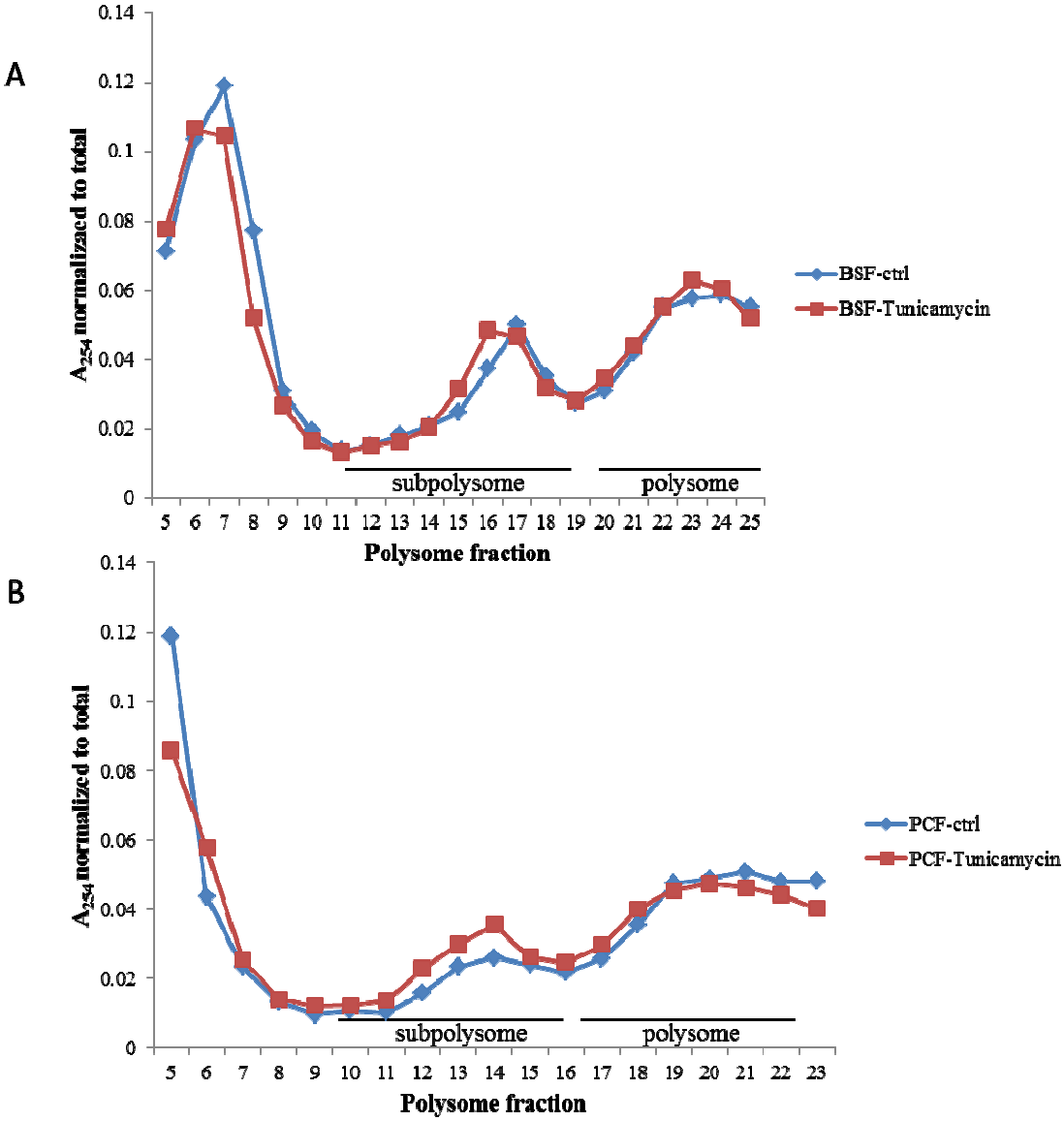


Figure 5.15: Representative absorbance (254 nm) traces of polysome profiles post fractionation and comparing (A) BSF and (B) PCF *T. brucei* that have been treated either with DMSO (ctrl) or tunicamycin.

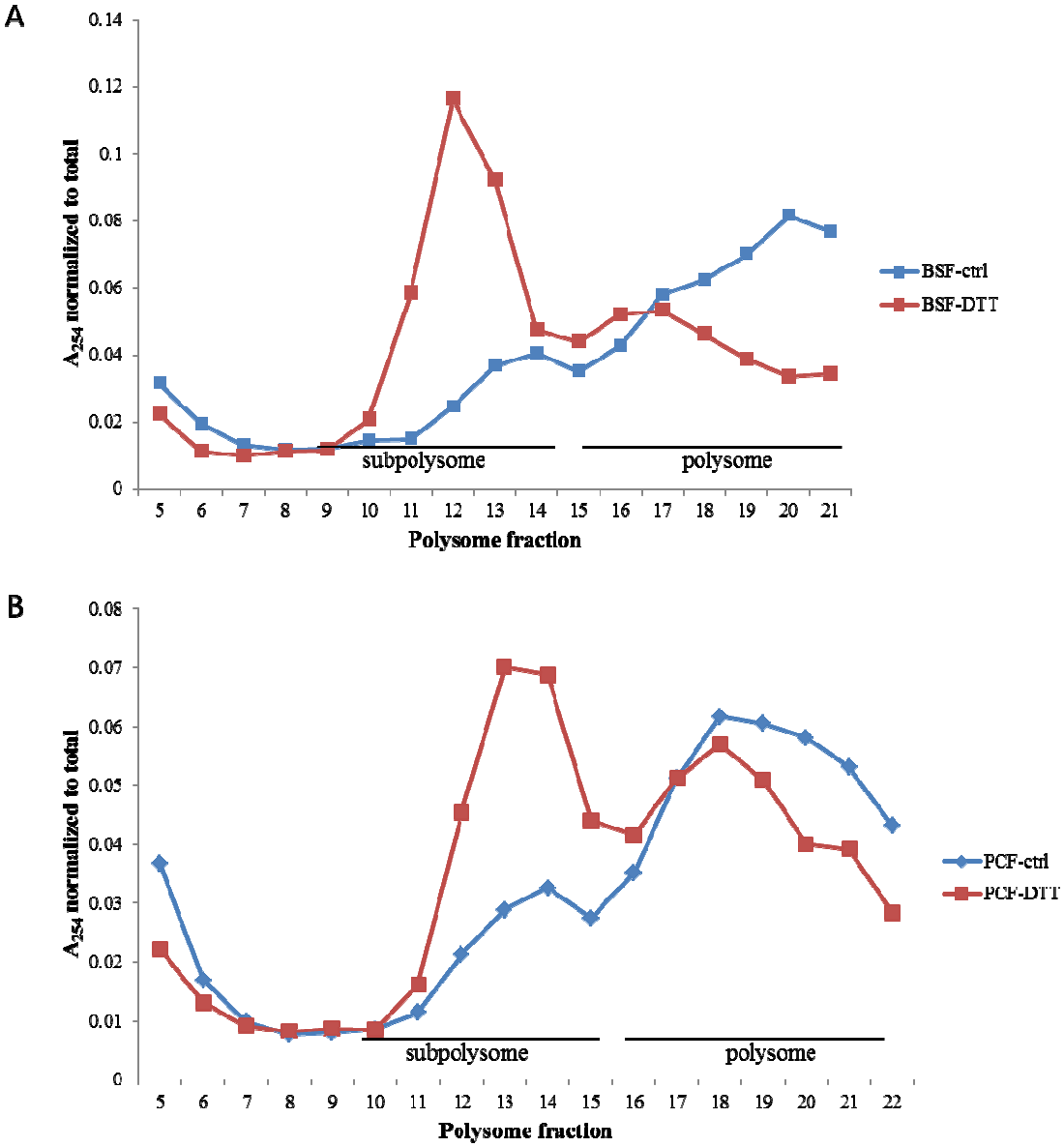


Figure 5.16: Representative absorbance (254 nm) traces of polysome profiles post fractionation comparing (A) BSF and (B) PCF *T. brucei* that have either been untreated (ctrl) or treated with DTT.

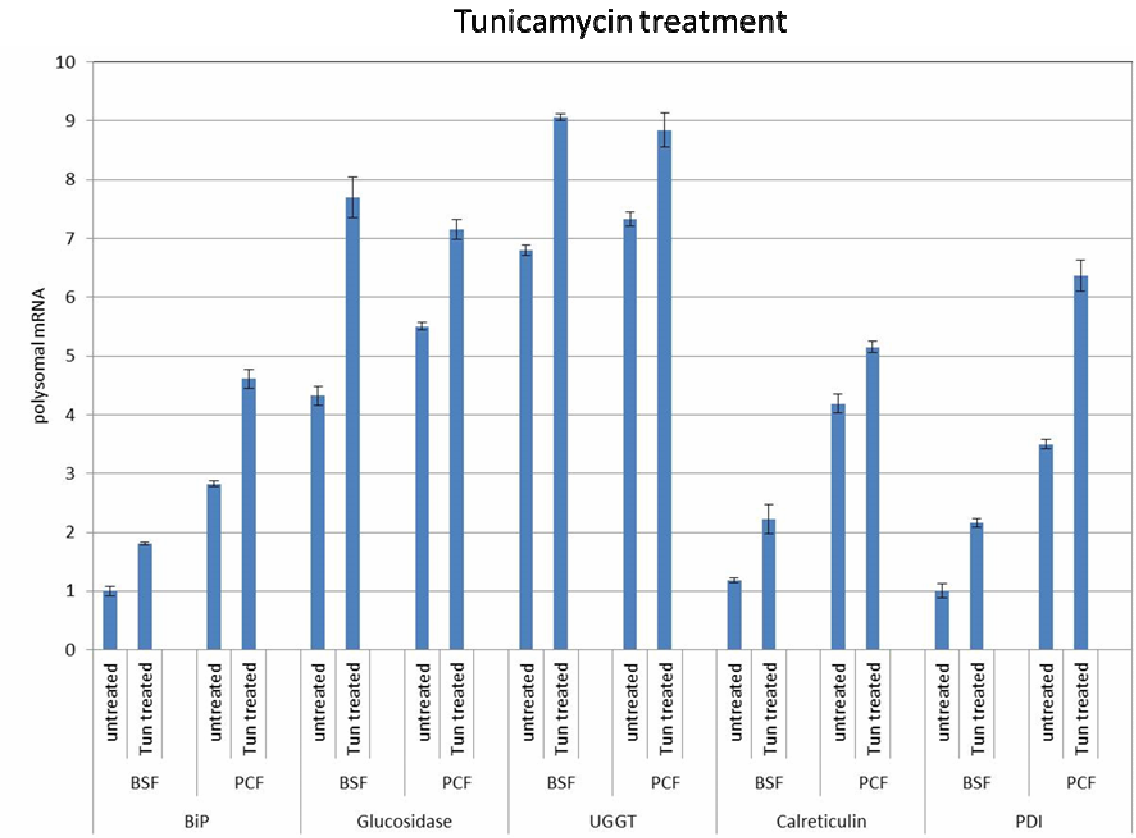


Figure 5.17: Polysomal enrichment of mRNAs. Total RNA from pooled polysomal fractions of BSF and PCF *T. brucei* following DMSO control (untreated) or tunicamycin treatment was purified and analysed by qRT-PCR for five candidate genes: BiP, glucosidase II, UGGT, calreticulin, and PDI to analyse their relative abundance in the polysomal fraction. Three biological replicates were done; the mean ± standard deviations are shown. Results were normalized to the constitutive gene: paraflagellar rod protein (PRP) (Brenndörfer and Boshart, 2010).

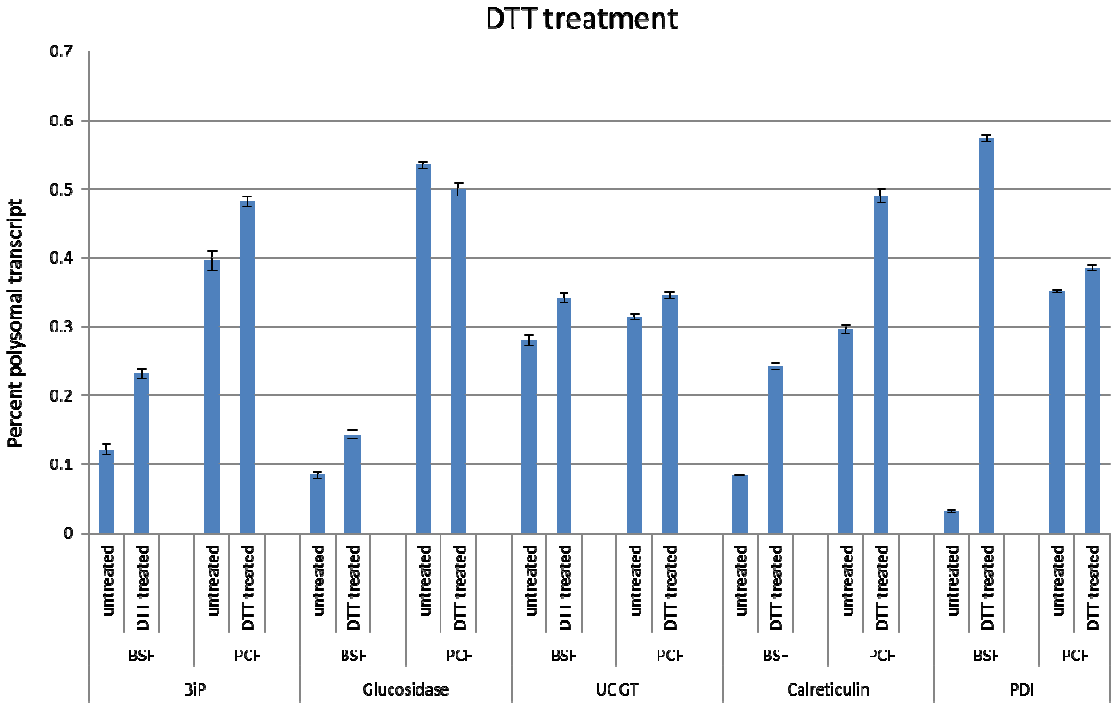


Figure 5.18: Polysomal enrichment of mRNAs. Total RNA from pooled subpolysomal and polysomal fractions of BSF and PCF *T. brucei* following control (untreated) or DTT treatment was purified and analysed by qRT-PCR for five candidate genes: BiP, glucosidase II, UGGT, calreticulin, and PDI to analyse their relative abundance in the polysomal fraction. Three biological replicates were done; the mean ± standard deviations are shown. Results were normalized to the constitutive gene: telomerase reverse transcriptase (TERT) (Brenndörfer and Boshart, 2010).

5.8 Comparative transcriptomics by RNA-seq of total and polyribosomal poly (A+) RNA purified from untreated or tunicamycin treated bloodstream form *T. brucei*

In order to gain insight into the widespread effects of ER stress on the BSF transcriptome and translatome, we performed RNA-seq on total mRNA and polysomal mRNA comparing tunicamycin treated to DMSO control treated bloodstream form *T. brucei*. Two biological and two technical replicates were sequenced using an IlluminaMiSeq sequencer. A high correlation for the read counts of biological replicates (average Pearson correlation coefficient R = 0.948) confirmed the reproducibility of the methodology (Figure 5.19). However, scatter plots comparing the two conditions examined (untreated (WT) and Tunicamycin (Tun) treated cells) did not reveal any differences between whole cell mRNA or in polysome associated mRNA transcript abundance (Figure 5.20). Furthermore, during the analysis, we also observed that the genes involved in the ER quality control pathway, in which we previously observed the "polysome under-representation", no longer displayed this phenotype in the untreated WT sample.

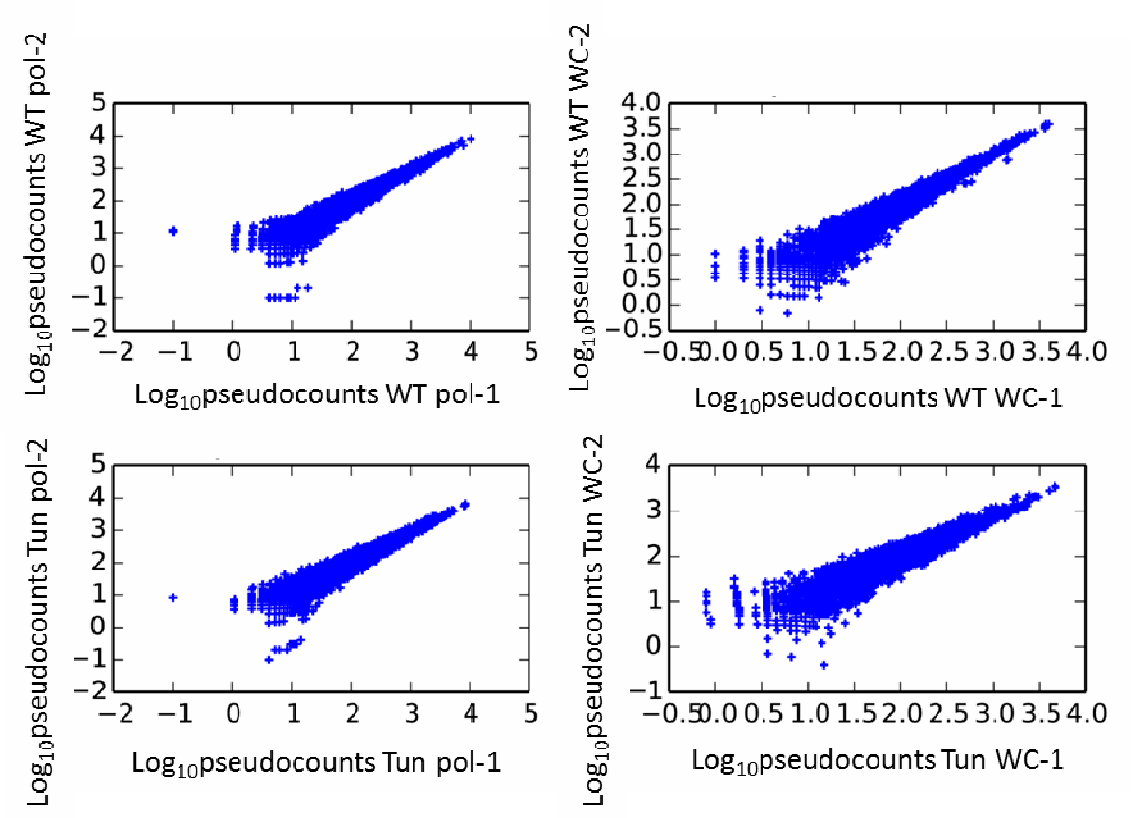


Figure 5.19: Scatter plot showing the comparison of biological replicates. BSF untreated (WT) and Tunicamycin treated (Tun) $\log_{10}$ of the normalized reads (pseudocounts) were compared between whole cell derived total mRNA (WC) and polysome associated (pol) transcripts for two biological replicates of RNA-seq data. Graph generated by Michele Tinti, MAJF lab.

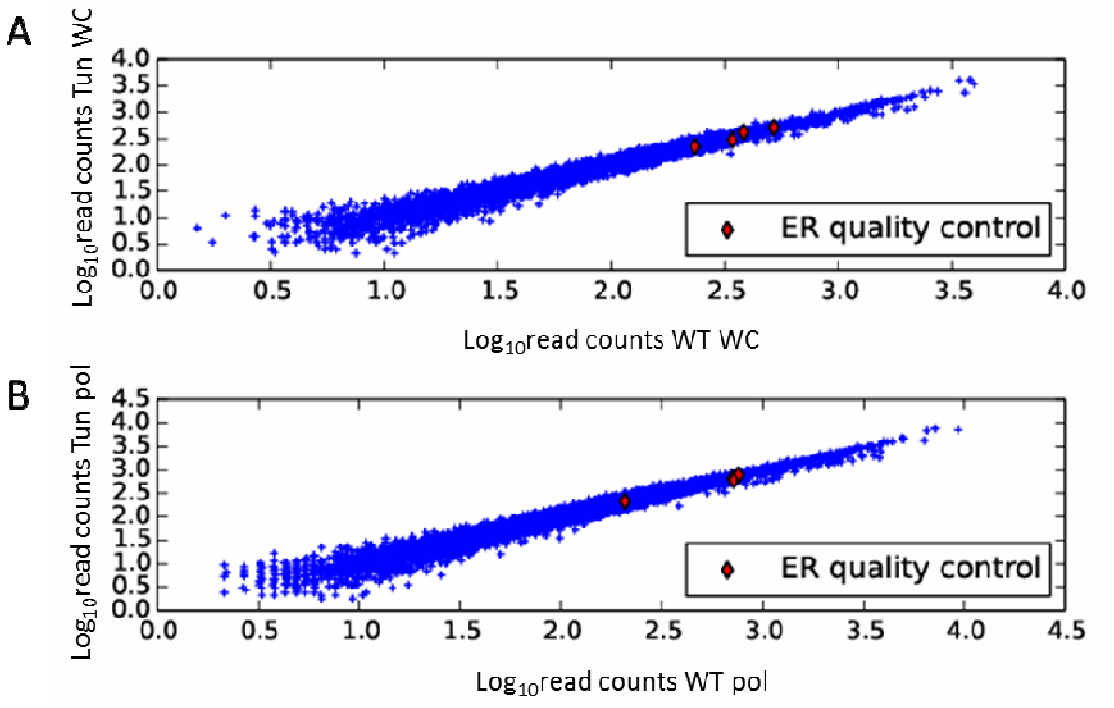
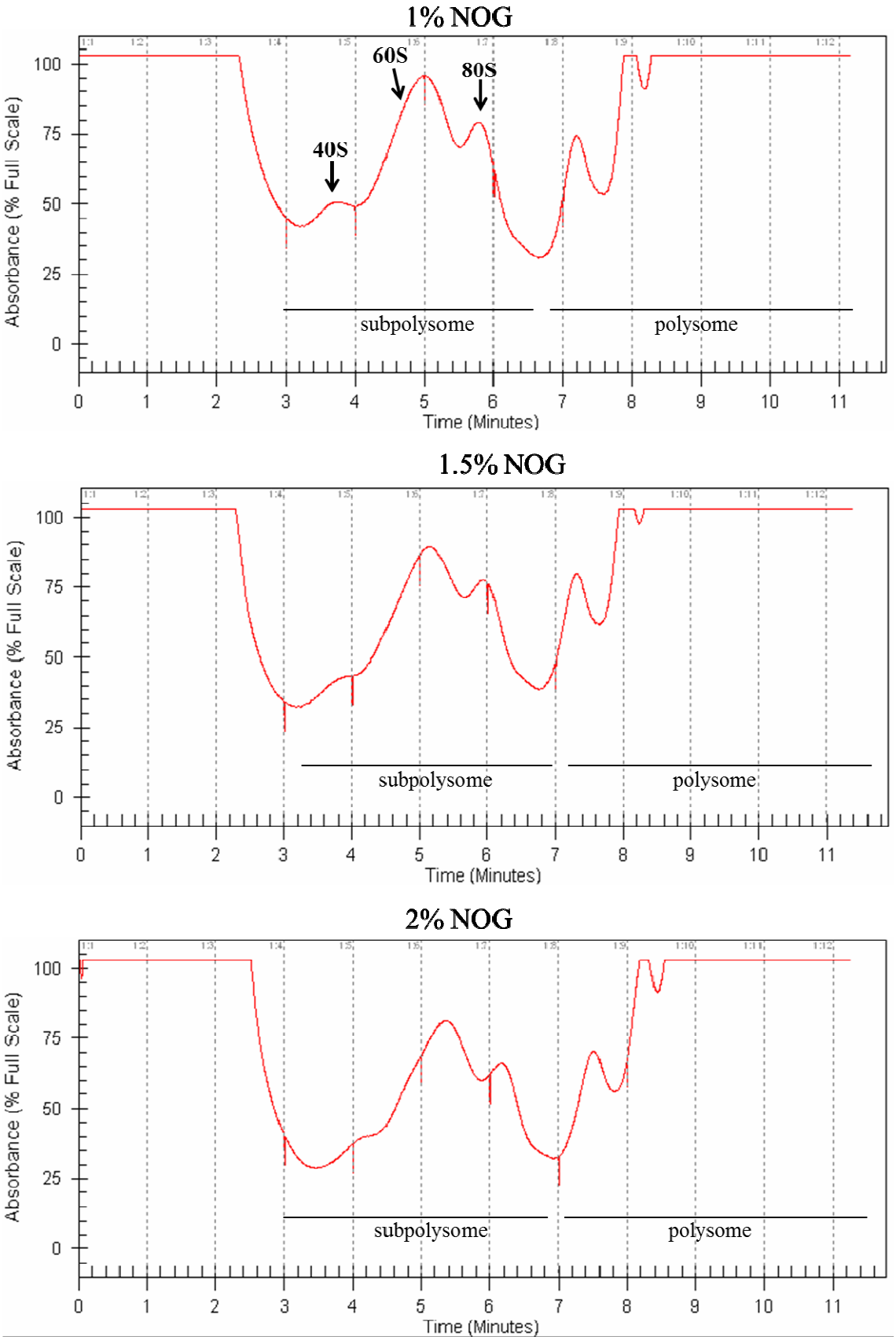


Figure 5.20: Scatter plot showing the comparison of normalized read counts. Correlation between untreated (WT) and Tunicamycin treated BSF *T. brucei* normalized read counts derived from (A) whole cell (WC) mRNA and (B) polysome associated (pol) mRNA. Genes involved in the ER quality control pathway which were analysed by qRT-PCR are designated as red diamonds. Graph generated by Michele Tinti, MAJF lab.

5.9 Effect of detergent concentration on the polysomal association of transcripts

Following the discrepancy in the results of the two RNA-seq experiments, the differences between the two experiment methodologies were narrowed down to the addition of DMSO to the untreated cells and the concentration of n-octyl-glucoside (NOG) detergent being increased from 1% to 2% in the lysis buffer. The addition of DMSO was discounted as the source of the discrepancy because the final concentration

of DMSO was only 0.25%. This lead to the hypothesis that under-representation of transcripts in the polysomal mRNA fraction might be an artefact of detergent concentration. To test this, polysome association of transcripts translated by cytosolic ribosomes was compared with signal peptide containing transcripts translated by ER-associated ribosomes, using a titration of NOG. BSF *T. brucei* lysates were prepared with 1-10% NOG lysis buffer and resolved through a sucrose gradient as described above. Polysome profile analysis did not indicate any differences in the distribution of ribosomal RNA across the gradient with increasing NOG concentration (Figure 5.21). The polysomal fractions from these polysome profiles were then pooled, purified, and compared to whole cell extracted transcript levels using qRT-PCR (Figure 5.22). If 1% NOG detergent was not enough to completely solubilize ER associated ribosomes, lysing the cells in higher concentrations of NOG should have increased the abundance of polysomal transcript, however, this was not observed (Figure 5.22, panel B).


1% NOG
60S
80S
40S
subpolysome
polysome
Absorbance (% Full Scale)
Time (Minutes)
1.5% NOG
subpolysome
polysome
Absorbance (% Full Scale)
Time (Minutes)
2% NOG
subpolysome
polysome
Absorbance (% Full Scale)
Time (Minutes)

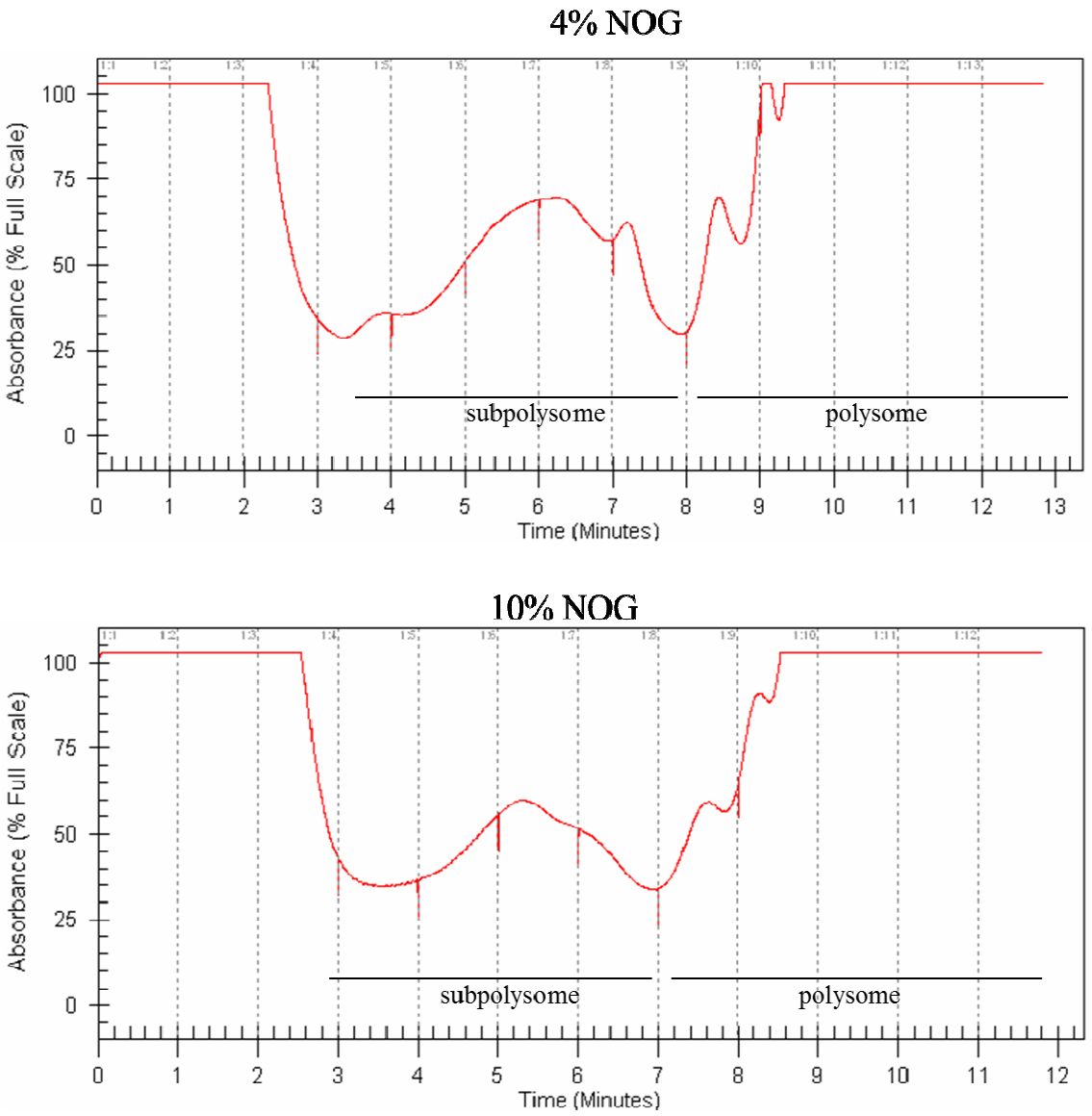
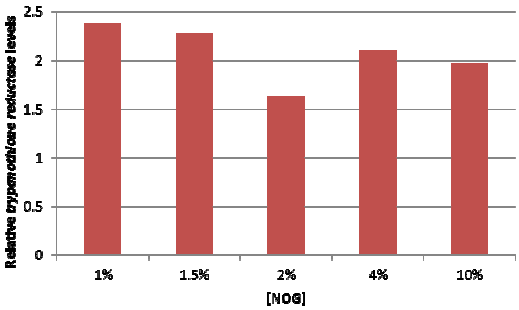
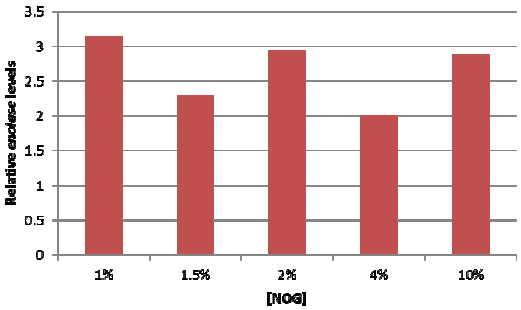
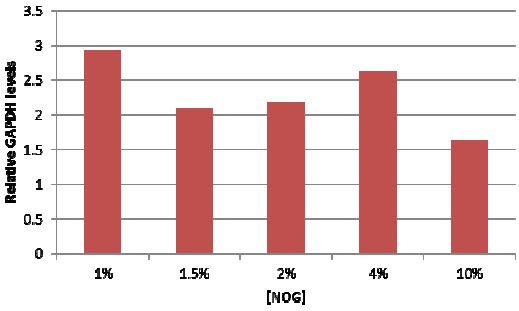


Figure 5.21: Representative absorbance traces of a BSF polysome profiles during collection of fractions. The first seven fractions contain subpolysomes and the last four fractions contain polysomes.Free monosomes (80S) and ribosomal subunits (40S and 60S) are indicated with arrows. Representative bioanalyserelectropherogram of purified RNA from the gradient depicting the three characteristic rRNA peaks. The baseline of polysome profile was normalized to the absorbance of a blank gradient.

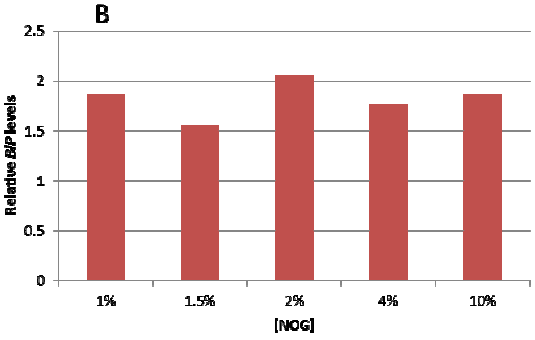
A

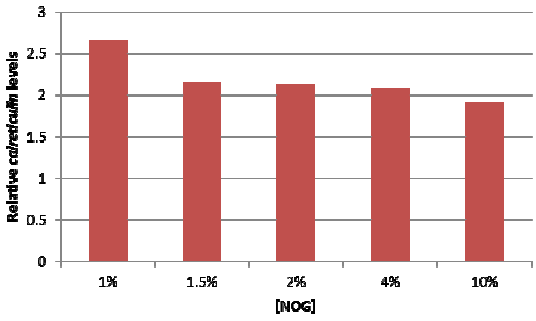

Relative cyp*meth/oxo reductase* levels
2.5
2
1.5
1
0.5
0
1%
1.5%
2%
4%
10%
[NOG]

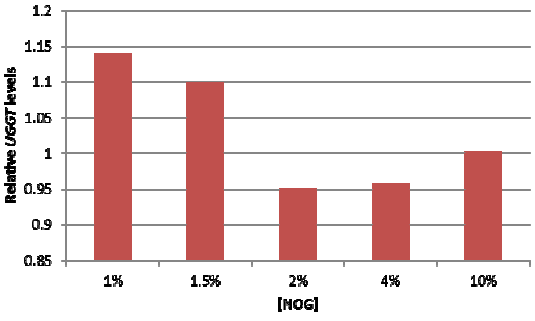

Relative *esotase* levels
3.5
3
2.5
2
1.5
1
0.5
0
1%
1.5%
2%
4%
10%
[NOG]

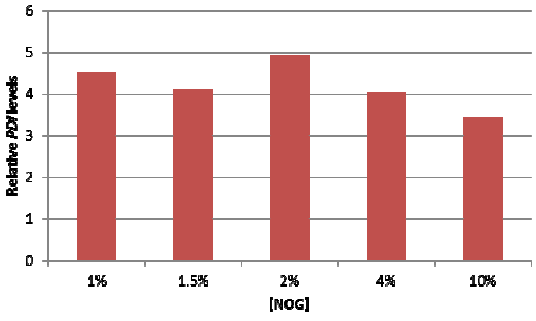

Relative *GAPDH* levels
3.5
3
2.5
2
1.5
1
0.5
0
1%
1.5%
2%
4%
10%
[NOG]

B


Relative *BiP* levels
2.5
2
1.5
1
0.5
0
1%
1.5%
2%
4%
10%
[NOG]


Relative *calreticulin* levels
3
2.5
2
1.5
1
0.5
0
1%
1.5%
2%
4%
10%
[NOG]


Relative *UGGT* levels
1.2
1.15
1.1
1.05
1
0.95
0.9
0.85
1%
1.5%
2%
4%
10%
[NOG]


Relative *PDI* levels
6
5
4
3
2
1
0
1%
1.5%
2%
4%
10%
[NOG]

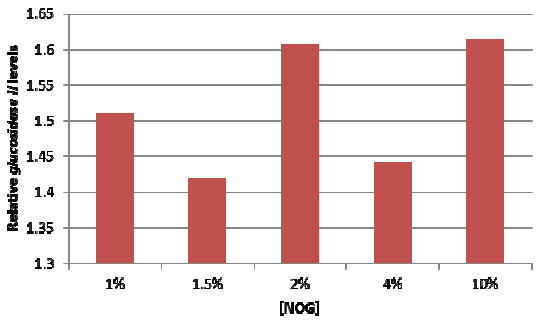

Relative *glucosidase II* levels
1.65
1.6
1.55
1.5
1.45
1.4
1.35
1.3
1%
1.5%
2%
4%
10%
[NOG]

Figure 5.22 (*previous page*): Distribution of polysomal transcripts with increasing detergent concentration. BSF *T. brucei* lysed with an increasing concentration of NOG detergent (1-10%) and RNA from whole cells and pooled polysomal fractions was purified and analysed by qRT-PCR. The log_2FC of polysomal transcript as a function of NOG concentration is shown (A) Three candidate genes that are translated by cytoplasmic ribosomes: typanothione reductase, enolase, GAPDH and (B) five candidate genes with signal peptides that are translated by ER associated ribosomes: BiP, calreticulin, UGGT, PDI, and glucosidase II were chosen to evaluate their relative abundance in the polysomal fraction. Results were normalized to the constitutive gene: telomerase reverse transcriptase (TERT) (Brenndörfer and Boshart, 2010).

5.10 Comparing the T. brucei transcriptome and proteome

In order to gain insight into the relationship between translational efficiency and protein abundance, the correlation between the RNA-seq data set with a stable isotope labelling by amino acids in cell culture (SILAC) proteomic data set was examined (Figure 5.23) (Urbaniak et al., 2013). Correlations were determined by comparing the Log_2 FC in whole cell derived (WC) mRNA, polysome associated mRNA or subpolysomal mRNA to protein. Surprisingly, the correlation between polysome associated transcripts and proteins was slightly lower than the correlation between total mRNA and protein (Figure 5.23, panels A and B). However, the very weak correlation between subpolysomal transcripts and proteins conformed to our expectations (Figure 5.23, panel C). The lack of improvement in polysome associated transcriptome correlation with protein abundance suggested that contrary to our initial hypothesis, using polysome profiling and the described data analysis pipeline, mRNA recruitment to ribosomes does not significantly contribute to steady state protein levels in *T. brucei*.

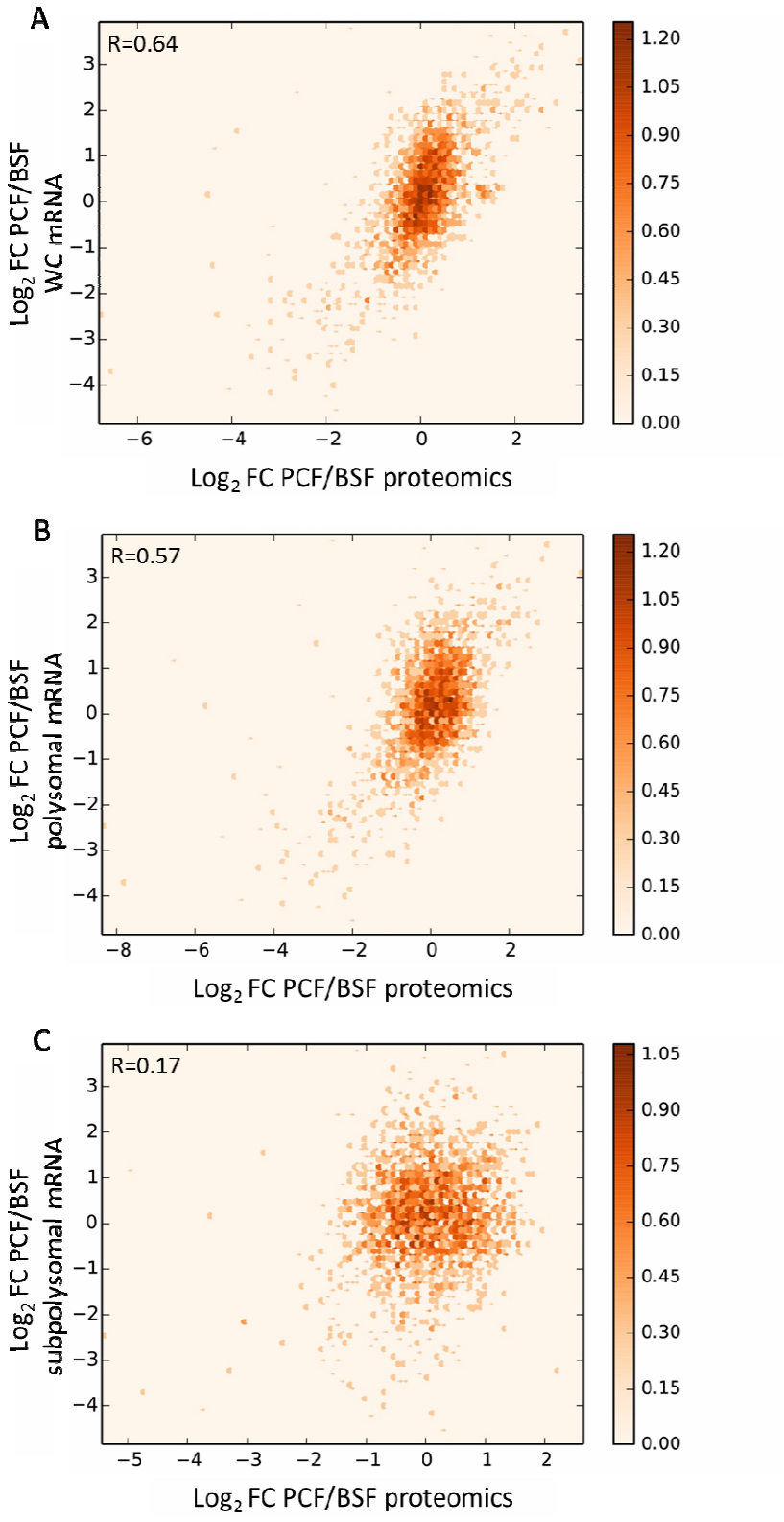

A
R=0.64
Log2 FC PCF/BSF WC mRNA
Log2 FC PCF/BSF proteomics
B
R=0.57
Log2 FC PCF/BSF polysomal mRNA
Log2 FC PCF/BSF proteomics
C
R=0.17
Log2 FC PCF/BSF subpolysomal mRNA
Log2 FC PCF/BSF proteomics

Figure 5.23 *(previous page)*: Hexagon binning histograms comparing transcriptome and proteome fold changes (FC). The $\log_2$ FC of PCF/BSF for (A) whole cell (WC) derived total mRNA transcripts, (B) polysomal mRNA transcripts and (C) subpolysomal mRNA transcripts were compared to the $\log_2$ FC of PCF/BSF SILAC proteome. In this histogram, the chart area is tessellated by a grid of hexagons and the number of points falling into each hexagon is plotted using a color ramp in proportion to the counts. The color ramp denotes the $\log_{10}$ of the number of points contained within the hexagon (i.e. dark orange points represent the most frequent correlation values). Graph generated by Michele Tinti, MAJF lab.

DISCUSSION

DISCUSSION

6.1 *Tb*CMT1 null and conditional null mutants

Kinetoplastids acquire their 5´mRNA cap through the trans-splicing of a 39nt SL RNA containing the unique cap4 structure (Bangs et al., 1992; Mair et al., 2000b). While the capping enzymes which form the cap0 have been biochemically characterized, the question of cap essentiality remains unanswered in bloodstream form *T. brucei.* Conditional null mutants of the monomeric methyltransferase *Tb*CMT1 with a tetracycline inducible ectopic copy of *Tb*CMT1 were generated in BSF trypanosomes. When cell growth was observed in non-permissive (-tet) conditions, the growth rate was nearly equivalent to cells grown in permissive (+tet) conditions. This suggested *Tb*CMT1 is not essential for parasite survival *in vitro* which was consistent with the results of *Tb*CMT1 RNAi knockdown in PCF *T. brucei*(Takagi et al., 2007).

As further and final validation that *Tb*CMT1 is not essential for parasite survival in culture, we generated *TbCMT1* null mutants (without an ectopic copy). Consistent with the results from the conditional null mutants, the growth rate of *Tb*CMT1 null cells did not display any defect, they appeared and behaved as normal. We next investigated whether *Tb*CMT1 influences *T. brucei* parasitemia in animal infections. Infections with SM cell lines grew well however, intriguingly, the parasitemia in *Tb*CMT1 null infected mice was very low (Figure 3.7). To understand if this *in vivo* effect was the direct result of *Tb*CMT1 loss or an off-target effect, mice infected with *Tb*CMT1 conditional null mutant cells were included in the study. The presence of doxycycline (in the drinking water of the mice) induced expression of the ectopic copy, and restored the growth of the parasites *in vivo*. This suggested the decrease in parasitemia observed in *Tb*CMT1 null infected mice was a direct result of the absence of *Tb*CMT1. Despite significant decrease in parasitemia of *Tb*CMT1 null infection compared to SM cell infection, the

incomplete eradication of parasites from mouse blood at the end of the 3-day study suggests *Tb*CMT1 may not be an essential for infectivity, but rather a virulence factor. The disparity between the *in vitro* and *in vivo* results is interesting and emphasizes the value of examining gene function with an *in vivo* animal model when possible. It appears gene essentiality is complicated by the parasites ability to tolerate the absence of a protein's function in culture only to reveal their capacity to infect is diminished in an animal model.

In pursuit of understanding the cellular functions affected by *Tb*CMT1 deletion, polysome profiling was used to compare translational activity between SM and *Tb*CMT1 null cells. Both cell lines were equally translationally active seen in the large proportion of RNA in the polysomal fractions (Figure 3.8). However, polysome profiles only offer an overall picture of cells translation, information about which proteins have changed their translational efficiency cannot be detected without analyzing individual transcripts. It would be interesting to explore which transcripts that are translationally up/down regulated in response to *Tb*CMT1 functional loss.

To gain more insight into the mechanisms underlying the decreased virulence of *Tb*CMT1 null mutants, we performed RNA-seq on SM and *Tb*CMT1 null mutant cell lines. With high reproducibility of the biological replicates, we detected 29 differentially expressed genes which exhibited a 2-fold or greater change ($P<0.05$) (Table 3.1). As RNA-seq becomes increasingly popular, several different data analysis processing pipelines have been developed. Here we used the most accepted differential expression packages (cuffdiff and DeSeq) with three different combinations of parameters and selected the genes that were common to all three modes of analysis for further study. Overall, the difference between SM and *Tb*CMT1 dKOtranscriptomes was minimal; the correlations between sample conditions are indistinguishable from

correlations between biological replicates. When the sequencing reads were mapped to the genome, there were 972 reads which mapped to unannotated exons. The average size of these novel (potential) genes is 1980bp, a reasonable size for coding sequences. Although no further investigation has been carried out on these sequences, four of them were identified as differentially expressed in the experiment. It would be interesting to examine these four sequences more closely for splice sites, regulatory motifs and protein expression in the available proteomic datasets for insight into their function.

Of the 21 genes which were found to be downregulated, the majority belong to the cysteine peptidase family (Table 3.1). Although the genomic sequences of the cysteine peptidases are nearly identical, making sequence mapping software distribute the reads arbitrarily among the 11 proteins, overall, the 'average' transcript abundance of this group of genes is downregulated in the absence of *Tb*CMT1. The results of the *in vivo* infectivity combined with RNA-seq point to an interesting possible role for *Tb*CMT1 in *T. brucei* because parasitic protozoa are known to have an abundance of cysteine peptidases that are important for survival of the parasite in mammals (Rosenthal, 1999). Further, studies in *Leishmaniamexicana* have seen cysteine peptidase deficient amastigotes infective *in vitro* but with impaired infectivity *in vivo* (Denise et al., 2003; Mottram et al., 2004). The relationship between *Tb*CMT1 and cysteine peptidases is unknown, and the results require additional validation to understand the scope of *Tb*CMT1 loss on the pathways involving cysteine peptidases.

The enzymes that were expected to be affected by the absence of *Tb*CMT1 were the other capping enzymes. However, neither the monomeric guanylyltransferase (*Tb*CE1) not the bifunctional enzyme (*Tb*CGM1) displayed changes in transcript abundance suggesting *Tb*CMT1 does not does not coordinate its function with the other capping enzymes *in vivo*. Given the extensive heterogeneity of SAS and PAS observed through

RNA-seq studies in *T. brucei,* it would be interesting to investigate changes in SAS and PAS site usage in *Tb*CMT1 null cells and explore the possibility that *Tb*CMT1 plays a role in alternative *trans*-splicing (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010).

6.2 *Tb*CGM1 conditional null mutants and RNAi

In addition to the capping enzymes characterized in other eukaryotes, trypanosomatids have an additional guanylyltransferase and methyltransferase in the form of a bifunctional enzyme (Takagi et al., 2007).The significance of the bifunctional capping enzyme remains unclear, so we proposed to study its essentiality and functionality BSF *T. brucei*. Despite several attempts, it was not possible to generate conditional null mutants of *Tb*CGM1. When initial attempts to replace the first allele with the PAC resistance cassette failed, a *Tb*CGM1 ectopic copy was introduced before replacement of the first allele in case of haploinsufficiency. Indeed, with the ectopic copy present, the first allele was successfully replaced, but the tetracycline regulation of the ectopic copy was lost and resulted in constitutive expression of the ectopic copy. It was thought that the loss of tetracycline regulation occurred by genetic rearrangement to prevent expression of the tetracycline repressor protein (Roper et al., 2002). In case ectopic expression from the pLew100 vector was not high enough to sustain gene replacement, an alternative ectopic copy expression vector (pLew100 v5) with 10-fold higher tet-induced expression was introduced. Finally, the first allele was replaced with the HYG resistance cassette, but when the *TbCGM1*$^{-/+}$, *TbCGM1*^{Ti} cells were cultured in non-permissive conditions the growth rate was not impaired, disproving the previous hypothesis that the cells are haploinsufficient for *Tb*CGM1. However, attempts to replace the second allele with the PAC resistance cassette were yet again met with misincorporation of the construct.

Unable to answer the question of essentiality using conditional null mutants, we changed the strategy to knockdown *Tb*CGM1using the inducible RNAi system (Alsford et al., 2005). Activation of the dsRNA targeting *TbCGM1* resulted in 67 ± 17% knockdown of *Tb*CGM1 mRNA after 24 h and cell death by 48 h, demonstrating the essentiality of *Tb*CGM1 in culture (Figure 3.14). This result is consistent with the results of *Tb*CGM1 RNAi knockdown in PCF *T. brucei,* although, in that case the cells survived considerably longer after induction (Ruan et al., 2007; Takagi et al., 2007). Further characterization of *Tb*CGM1 function was not carried out in this project, but Takagi and colleagues demonstrated that knockdown of *Tb*CGM1 resulted in the accumulation of uncapped SL RNA in PCF *T. brucei*. With the unique capacity for both guanylyltransferase and methyltransferase activity, a role for *Tb*CGM1 in capping was expected and confirmed by their work.

The monomeric methyltransferase, *Tb*CE1, did not exhibit any effects of cell growth or SL RNA capping when subject to knockdown by RNAi in PCF *T. brucei*(Ruan et al., 2007; Takagi et al., 2007). It is possible that the residual transcript from the knockdown is sufficient for cell survival in culture and a conditional null mutant would be necessary to validate its essentiality. Taken together, the fact that neither *Tb*CE1 nor *Tb*CMT1 could rescue the growth defect of *Tb*CGM1 knockdown suggests their roles do not function in the same pathways as *Tb*CGM1.

In the future, to study the essentiality and potential redundancy of the methyltransferase domain of the bifunctional enzyme *Tb*CGM1, an RNAi ‘readout’ system will be established. We hypothesize that the *Tb*CGM1-depletion-induced phenotype (cell death) could be rescued by the presence of a constitutively active *Tb*CGM1 wobble mutant which is resistance to depletion by RNAi. Assuming this is the case, we could then dissect which domain of the bifunctional enzyme is essential, by introducing

catalytically inactive methyltransferase or guanylyltransferase wobble mutants and assessing parasite viability following *Tb*CGM1depletion by RNAi.

In summary, while there are still many unanswered questions about the specific roles of the modified cap nucleotides and the enzymes that catalyse their formation, we suggest that the mRNA 5′ cap structure is essential in BSF *T. brucei* and that the *Tb*CGM1methyltransferase and/or guanylyltransferase catalytic domains may be a drug targets against kinetoplastids.

6.3 TFA hydrolysis of mRNA cap structure analogues

To investigate the possibility that some of the mRNA 5′ cap nucleosides in *T. brucei* are structurally different from conventional 7-methylguanosine, a method was developed to isolate the cap from endogenous mRNA and analyse the resulting structure by mass spectrometry. The possibility of novel mRNA cap structure(s) was suggested by the existence of an additional guanylyltransferase/methyltransferase (*Tb*CGM1) and the unusual cap4 structure that is found uniquely in kinetoplastids. Our analytical approach was based on the extreme acid liability of pyrophosphate bonds which allowed us to differentiate between the triphosphate bond linking the 5′ cap to the transcript, and all of the other bonds connecting individual nucleotides in a transcript (Figure 4.1). The initial experiments showed that it was possible to resolve the starting material from the theoretical hydrolysis products by mass spectrometry (Figure 4.5). Next, the specificity of TFA hydrolysis on pyrophosphate bonds was tested on the theoretical hydrolysis products, and as expected, structures which lack pyrophosphate bonds (G and GMP) are resistant to hydrolysis (apart from the N-glycosidic bond between the ribose and the base). To explore the chemical properties of the cap, the commercially available analogue m7GpppG which mimics the 5′RNA cap structure joined to the first transcribed nucleotide, was used to optimize the methodology. Initial optimization of

TFA hydrolysis produced robust peaks of both expected hydrolysis products (GDP and m7GDP) and served as a proof of principle for the use of mild acid to cleave the 5′cap from an RNA molecule (Figure 4.7). TFA hydrolysis also yielded an undesirable ribose-diphosphate degradation product which resulted from the cleavage of the N-glycosidicbond between the ribose sugar and base. The presence of this degradation product indicated that it was necessary to optimize TFA hydrolysis such that the yield of desired products was maximized but not pushed to the point of excessive degradation; this balance was achieved by 20 min hydrolysis.

With the exception of the 5′ cap, which is connected to the first nucleotide by three phosphate groups, each nucleotide in an RNA molecule is connected to the adjacent nucleotide by a single phosphate group. We sought to exploit this difference by TFA hydrolysis because the cap is the only structure in mRNA that can be released with a pyrophosphate group. In precursor ion scanning mass spectrometry, the specific loss of an ion (in this case pyrophosphate) from the precursor molecule is monitored across the spectrum. This highly selective filter enables the identification of pyrophosphate containing molecules amongst a complex mixture and therefore potential novel cap structures of any mass can be identified. Combining TFA hydrolysis with the precursor ion scanning method, analysis of the cap analogue m7GpppG yielded both predicted precursor ions: $[m7GDP]^-$ and $[GDP]^-$ (Figure 4.11). It's important to note that there would only be one predicted precursor ion, $[m7GDP]^-$ in an mRNA sample because the first nucleotide would remain attached to the rest of the mRNA. However, because the cap analogue m7GpppG only has a single 'nucleotide' on its 3′end, TFA hydrolysis generated two precursor ions. Despite a predicted equal ratio of GDP to m7GDP after hydrolysis, the intensity of GDP in the spectrum appears greater because the methylation at the N7 position confers a positive charge to the cap structure, thereby reducing its ionization in the negative ion mode of the mass spectrometer (Figure 4.11).

Nonetheless, qualitatively, observation of the $[m7GDP]^-$ ion indicated that m7GDP was effectively cleaved from the cap analogue by TFA hydrolysis and detected by precursor ion mass spectrometry.

6.4 Precursor ion scanning method optimization

The method was then taken forward for testing on *in vitro* transcribed RNA to bridge the gap between the simplicity and purity of the commercially available cap analogue and the complexity of cell derived RNA. Despite several attempts to generate m7G capped *in vitro* RNA, the precursor ion could not be detected. While the benefit of an *in vitro* system is the ability to quickly and easily generate large quantities of *in vitro* transcript, the amount of cleaved m7GDP could not be determined with the same certainty as the analogue and therefore may have been below the level of detection of the mass spectrometer. To optimize the sensitivity of the mass spectrometer, titration of a GDP standard was used as a surrogate for the cap structure. A large improvement in sensitivity was achieved when tributylammonium acetate (TBAA) replaced triethylammonium acetate (TEAA) as the ion-pairing agent. As a stronger ion-pairing agent, TBAA bound the cap structures with a higher affinity creating better chromatographic separation and controlled elution. A sensitivity of 10 pmol was achieved with the TBAA mobile phase buffer, a significant improvement on the initial sensitivity of 1 nmol.

Encouraged by this result, the method was tested on total RNA extracted from mammalian HEK-293 cells. Although the precursor ion scanning method is a very selective filter, the complexity of the biological sample revealed that the signal to noise ratio of the spectra was too low to observe the mRNA 5´cap diphosphate ion, m7GDP. The background of precursor ions that contain an *m/z* 159 (pyrophosphate) fragment overwhelmed the potential cap derived precursor ions. Increasing the amount of input

RNA would not improve the spectrum as the number of unidentifiable precursor ions would also increase proportionally. Therefore, it was necessary to enrich the sample prior to hydrolysis and analysis by mass spectrometry by extracting the cap-containing mRNA out of the mixture of tRNA and highly abundant rRNA.

6.5 Optimization of mRNA extraction buffer

The poly(A) tail can be used to efficiently recover mRNA from the other RNA molecules using oligo-dT cellulose. In order to use this technique in conjunction with mass spectrometry, it was necessary to identify a suitable substitute for the standard oligo-dT buffers because salts and detergents cannot be used in mass spectrometers. At the same time, further purification to exchange the mRNA into a buffer that is compatible with mass spectrometry will cause sample loss. Salts overload the system with charged salt ions and generate multiple salt adducts of the desired analyte which can saturate the detector. Therefore, components of the oligo-dT cellulose binding buffer were methodically removed and replaced with salts that can be better tolerated by the system. To explore the effects of binding buffer modification on mRNA extraction, total RNA from HEK-293 cells was enriched for mRNA using five different buffers and analysed for rRNA contamination on an Agilent Bioanalyser (Figure 4.13). Detergents interfere with the evaporation process on the electrospray of the mass spectrometer, and therefore the first component removed from the olig-dT buffer was SDS which resulted in a less efficient oligo-dT binding buffer. Decrease in the concentration of EDTA also resulted in a less efficient oligo-dT binding buffer. However, removal of SDS, Tris together with the substitution of NaClby sodium acetate resulted in a binding buffer that is both effective for mRNA enrichment and compatible with mass spectrometry.

6.6 Assessment of optimized methodology

Combing these improvements, the method was re-designed and HEK-293 cells were used to assess the complete method from the culture flask to the precursor ion scan (Figure 4.15). A peak at *m/z* 456 corresponding to m7GDP precursor ion was observed at the elution time previously seen for the m7GDP ion generated by hydrolysis of the cap analogue m7GpppG. The detection of this peak was encouraging for the progress of the method as it supports the seminal work of (Furuichi et al., 1975; Shatkin, 1976). However, the signal to noise ratio is still very low, evidenced by the unknown peak adjacent with nearly the same intensity as the m7GDP ion peak (Figure 4.15, panel A). Additionally, there appears to be a large peak corresponding to GDP ion (Figure 4.15, panels C and D). It is not possible to distinguish whether the GDP ion is derived from unmethylated mRNA cap or from the partial hydrolysis of GTP that could have been a contaminant of the mRNA population. In order to determine the source of the GDP ion, the effect of recombinant mammalian cap methyltransferase (RNMT) added to mRNA preparations could be examined. If treatment with RNMT increases the proportion of m7GDP to GDP ions, it would suggest the former explanation and it would then be interesting to study the quantitative distribution of capped vs methylcapped mRNA (Cowling and Cole, 2010). Peaks potentially corresponding to ADP and CDP were also detected, although appropriate standards would have to be examined to accurately assign their presence in the sample. The next experiment proposed is to analyse BSF *T. brucei* mRNA with this method and investigate the potential differences in their TFA hydrolysis products. Further optimization will be necessary to improve the signal to noise ratio, this could be done with time course of TFA hydrolysis using mRNA instead of the cap analogue to determine conditions where hydrolysis side effects are at a minimum. Once the method is robust it could be used to explore the cap structures of the *Tb*CMT1 null cell line and other kinetoplastids.

Although studies on the mRNA cap structure in *T. brucei* using thin-layer chromatography (TLC) identified 7-methylguanosine as the 5′ cap structure, spots that could not be identified with the available standards were also detected (Freistadt et al., 1987; Sutton and Boothroyd, 1988). Advances in the sensitivity of analytical analysis afforded by mass spectrometry allow us to revisit the original discoveries of mRNA cap structure and build upon their findings.

6.7 Preparation of polysomal RNA

Given the disparity in studies of changes in the transcriptome vs proteome between BSF and PCF *T. brucei*, our hypothesis was that mRNA recruitment to ribosomes might contribute to the regulation of gene expression (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010; Urbaniak et al., 2012). To test this hypothesis, polysome profile analysis was used to monitor the distribution of polysomal transcript between bloodstream form and procyclic form *T. brucei*. Polysome profiling is based on the principle that the high protein content of polysomes allows them to be separated throughout a sucrose gradient according to the number of ribosomes attached to the mRNA. Further, a particular mRNA's higher or lower than average association with ribosomes indicates potential involvement of gene-specific regulatory mechanisms (Pradet-Balade et al., 2001).

Initial polysome profile analysis revealed that both BSF and PCF trypanosomes appeared to be highly translating cells, indicated by observation that the majority of RNA is the polysomal fraction of the gradient. PCF cells also showed a noticeably greater ratio of polysomal to subpolysomal RNA than BSF cells confirming the higher rate of translation other studies have also observed in PCF cells (Vasquez et al., 2014).

To understand if ribosome association plays a role in the regulation of protein expression, the distribution of the transcripts for six candidate genes was analysed

across the polysome profile. All of the genes examined, with the exception of Aconitase, displayed a polysome profile distribution consistent with the published results of their life cycle specific expression (Figure 5.3). The control gene, DPMS, is required in abundance in both life cycle stages and therefore the majority of the transcript was in the polysomal RNA fractions. Although RAB18, a BSF specific protein, had a greater proportion of its transcripts in the BSF polysome fraction relative to the PCF polysome fraction, the effect was not profound and is unlikely to account for the BSF-specific nature of RAB18 protein expression. Transketolase, mMDH and CAP5.5 all clearly display a enrichment in PCF polysomal fractions which was expected as they have been shown to be procyclic specific genes. As a first approximation, the data for this small number of transcripts appeared mostly promising, however it would be difficult to establish trends for larger subsets of genes based on these methods and therefore a genome-wide approach using RNA-seq was pursued.

6.8 cDNA library preparation and preliminary RNA-seq

To prepare samples for RNA-seq analysis, cDNA libraries were generated from both total mRNA and polysome-associated mRNA transcripts. Following rRNA depletion by selection of poly(A) mRNA using oligo-dT magnetic beads, the mRNA was sheared into smaller ~200 nucleotide fragments for optimal sequencing coverage of smaller transcripts and to circumvent the potential bias of reverse transcriptase and to maximize coverage of the sequencer's paired-end reads. In addition to the potential sequence bias by reverse transcriptase enzyme, RNA secondary structure compounds the sequence bias in the cDNA library. However, both of these potential problems were alleviated by fragmentation of the mRNA prior to reverse transcription (Hansen et al., 2010; Mortazavi et al., 2008). It is important to note that unlike the established cDNA library protocol, our procedure enabled the library to be completed without PCR amplification, therefore eliminating the sample bias associated with variable amplification. The

procedure for the generation of cDNA libraries has been successfully optimized such that the amount of library necessary for Illumina sequencing can be generated without a large amount of initial RNA. This enables the removal of the PCR amplification step in favour of more reliable relative quantification of the dataset. Throughout the cDNA library preparation procedure, samples were monitored by Agilent Bioanalyser to ensure RNA integrity and cDNA quantity (Figure 5.6). Due to the ubiquitous presence of RNA exonucleases, the Bioanalyser is a useful tool for real-time quality control of the samples as well as troubleshooting difficulties with library generation. To begin to investigate the hypothesis that mRNA selection might contribute to the regulation of gene expression, two biological replicates of BSF and PCF polysomal and total (whole-cell) derived mRNA) were constructed for a preliminary RNA-seq experiment (Figure 5.7). The high correlations for the read counts of the biological replicated showed confidence in the reproducibility of polysome profile analysis and RNA-seq sample preparation. A matrix of scatter plots comparing all 8 samples simultaneously visualized the strong correlation between biological replicates contrasted with the weaker correlation between total mRNA and polysomal mRNA samples due to the differentially expressed genes. Indeed, hundreds of differentially expressed genes were identified, encouraging a more thorough analysis including the subpolysomal fraction, additional replicates, and retrieving a deeper transcriptome by performing the sequencing on an IlluminaHiSeq instrument.

6.9 RNA-seq of total, subpolysomal, and polyribosomal poly (A+) RNA purified from bloodstream form and procyclic form *T. brucei*

Three biological and three technical replicates of total, subpolysomal, and polysomal mRNA were sequenced andpairwise comparisons of the biological replicates showed strong correlations indicating they are reliable for analysis (Table 5.2). The stark differences between bloodstream and procyclic form *T. brucei* have been studied for

decades and recently shown on a global transcriptomic scale. Consistent with these findings, our dataset also detected thousands of significantly differentially expressed genes between the two life cycle stages (Table 5.1). Considering the contrasting environments these two forms of the parasite can survive and their different systems of immune evasion it is reasonable that so many genes change in transcript abundance.

The volume of data generated by genome wide studies such as RNA-seq are difficult to visualize because there are millions of reads and thousands of genes to analyse. To facilitate visualizing the relationships between samples we use software to represent the information in a two-dimensional format that can be easily interpreted. A principle component analysis (PCA) was performed to visualize the distinct groups of reads produced by the different samples (Figure 5.8). The PCA confirmed the reproducibility of the replicates with one exception, PCF-sub3. This sample contained the greatest deviation from other two biological replicates. Further analysis of this replicate revealed it had the lowest Pearson correlations with an average R= 97.7 compared to an average R= 99.8 for the BSF-wc samples, the best correlated replicates. There were 375 genes in sample PCFsub-3that are potential outliers. One possibility is that slight variation in the efficiency with which RNAs other than mRNAs are excluded during cDNA library preparation could exaggerate differences. This may have been because the subpolysomal fraction contained all of the free monosomes. Contaminating rRNAwould easily show up in such comparisons between replicate samples, however, following statistical analysis of the deviation caused by the potential outliers, it was decided to keep the PCF-sub3 sample as part of the dataset and any genes that are chosen for validation can be manually cross-referenced with the list outlier genes.

We then used this data set to ask question if translational efficiency might be regulated in a life cycle-specific manner. The strong correlation between BSF: PCF log-fold changes for total and poly suggests that the distribution of transcript abundance is the

same between total mRNA and polysome-associated mRNA (Figure 5.9). As rapidly proliferating and highly translationally active cells, *T. brucei* appear to have few transcripts in the cell that are not substantially polysome associated. That is not to say that few transcripts maintain distinct life-cycle differences, but rather that the differences are observed to the same extent in both in total mRNA and polysomal mRNA. Consistent with other transcriptomic studies, the largest upregulation of transcripts in the BSF were VSG and glucose transporter 1B (THT1) and the largest upregulation of transcripts in the PCF were procyclins(Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010). This result was expected as VSGs and procyclins are highly expressed in their respective life stages. Bloodstream form *T. brucei* are entirely dependent on glycolysis for energy and therefore the large increase in translation efficiency of the low-affinity, high capacity THT1 glucose transporter was also expected (Bringaud and Baltz, 1993).

A GO term analysis was performed to identify functional categories of genes that were up or downregulated between the lifecycle stages (Figure 5.10, panel B). The analysis produced very broad results of proteins in almost every category of cellular function, making it difficult to draw conclusions. However, the absence of differential expression between total mRNA and polysomal mRNA suggest translational efficiency does not significantly contribute to the regulation of gene expression in *T. brucei*. This conclusion contradicts the results observed by Vasquez and colleagues (Vasquez et al., 2014). Also interested in translational control of the gene expression in *T. brucei*, they performed ribosome profiling on BSF and PCF cells. Ribosome profiling is a higher resolution technique to polysome profiling, however the methodologies should produce similar results (Ingolia et al., 2012). Although they also describe a positive correlation between the translational efficiencies in the two life-cycle stages, a 117-fold range in translational efficiency in PCF and a 64-fold range in BSF was reported (Vasquez et al.,

2014). However, further analysis of the differences of sample preparation and data analysis would be necessary to make direct comparisons between the studies.

Following the result of the strong correlation between total mRNA and polysomal mRNA, it was unexpected to find transcripts that were under-represented on polysomes: i.e. the level of transcript in the total mRNA sample was greater than the level of transcript in the polysomal fraction.Because the former result suggested very little mRNA is untranslated, we were intrigued by the genes that did not follow the pattern and noticed that these genes are involved in the ER quality control pathway. This observation lead to the hypothesis that trypanosomes possess a reserve capacity for polysome recruitment in situations requiring a rapid response, such as ER stress. Two different stress-inducing compounds were added to *T. brucei* and the effect on polysomal transcript distribution was analysed. When treated with tunicamyicn, an N-linked glycosylation inhibitor, no overall change in the polysomal profile was detected (Figure 5.15) However, when the genes involved in the ER quality control pathway were analysed specifically, a small but robust increase in polysomal transcript was observed (Figure 5.17). In contrast, treatment with DTT resulted in a substantial shift of transcripts into the un-translated subpolysomal fraction. While both chemically induce stress, the effect of tunicamycin appears to be localized to the ER while DTT appears to ubiquitously affect the translatome (Figure 5.16). Analysis of the genes involved in the ER quality control pathway following DTT treatment showed similarly small and robust increases in polysomal transcript with the exception of PDI which displayed a large shift toward polysomal transcript (Figure 5.18). As DTT treatment disrupts disulphide bonds present in the cell, it might be expected for the cell to respond with a large increase in PDI to re-form the disordered bonds.

6.10 RNA-seqanalysis of tunicamycin treated vs untreated bloodstream form *T. brucei*

In order to gain insight into the widespread effects of ER stress on the BSF transcriptome and translatome, we performed RNA-seq comparing tunicamycin treated to untreated bloodstream form *T. brucei*.Surprisingly, a comparison of thepolysomal association of transcripts in the presence or absence of tunicamycin did not reveal any differences (Figure 5.20). This suggested that contrary to the hypothesis, inducing ER stress in *T. brucei* via treatment with Tunicamycin does not promote increased translation of ER quality control genes by higher association with polysomes. However, analysis of the under-represented genes studied individually by qRT-PCR showed that the "polysome under-representation" phenotype was not reproducible in the untreated RNA-seq sample. We scrutinized the methodology to try to identify the source of the difference between the experiments, however only two differences emerged. First, to control for solvent effects, DMSO was added to the untreated BSF cells at the same concentration and for equal duration as tunicamycin. However, the final concentration of DMSO was determined to be only 0.25% and therefore deemed inconsequential. The second difference was the use of 2% NOG detergent instead of the 1% NOG detergent used in the previous RNA-seq experiments. The concentration of detergent was increased in the lysis buffer to be sure that all rough ER membrane bound polysomes were solubilized and their translating mRNA purified. This lead to the hypothesis that under-representation of transcripts in the polysomal mRNA fraction might be an artefact of detergent concentration. In the event that detergent concentration affected the proportion of polysomal transcript, it was though that genes containing signal peptides were more likely to be subject to this potential artefact. This was investigated by the comparison of Trypanothione reductase, enolase, and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) which are all enzymes without signal peptides and are therefore presumed to be translated on cytoplasmic ribosomes (Aslett et al., 2010).

Conversely, the genes that appeared under-represented in the polysomal fraction (BiP, calreticulin, UGGT, PDI, and glucosidase II) contain signal peptides are therefore presumed to be translated on ER associated ribosomes (Aslett et al., 2010). Using increasing concentration of the NOG detergent, this selected group of genes were analysed for their proportion of polysomal transcript. However, increasing the concentration of NOG did not significantly increase the proportion of polysomal transcript in either set of candidate genes examined (Figure 5.22). Although this experiment was only performed once and requires additional replicates for statistical significance, there was no observable trend to support the idea that polysome association of transcripts with signal peptides is a function of detergent concentration. This suggested the inconsistency of transcript under-representation on polysomes might not be an artefact of detergent concentration. The procedure for cell lysis does not appear to be standardized because while some studies performing polysome profiling use 1% NOG, other studies used 1% Triton X-100 and still others lysed *T. brucei* through the sonication of frozen cells or vortexing with glass beads (Capewell et al., 2013; Djikeng et al., 2003; Smith et al., 2009; Vasquez et al., 2014). One possible reason for the discrepancy between the experiments is that the untreated cells were also under stress and therefore increased the polysomal association of stress-response transcripts. For future experiments it is important confirm the effectiveness of tunicamycin on the inhibition of glycosylation as well as to test untreated cells for the upregulation of stress factors to ensure the fidelity of the conclusions that will be drawn from the experiment.

6.11 Comparing the *T. brucei* transcriptome and proteome

Following the results of this study that the majority of transcripts are similarly distributed between total mRNA and polysomal mRNA fractions, combined with the results of other transcriptomic studies which found only 5.6-32% of the transcriptome

changes between BSF and PCF *T. brucei*, we endeavoured to compare translational efficiency with protein abundance (Figure 5.23) (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010). Correlations between the RNA-seq data set and (SILAC) proteomic data revealed the correlation between polysome associated transcripts and proteins was in fact slightly lower than the correlation between total mRNA and protein (Urbaniak et al., 2013). However, the very weak correlation between subpolysomal transcripts and proteins conformed to our expectations. Previously described correlations between SILAC proteomics data and three different microarray transcriptomic datasets showed higher correlations with two of the studies (R=0.83/0.86) and significantly lower correlation with the third study (R=0.2), illustrating the difficulty of combining data from different genomic studies (Urbaniak et al., 2012). Even so, the lack of improvement in polysome associated transcriptome correlation with protein abundance suggeststhat our initial hypothesis that mRNA recruitment to ribosomes might contribute to the regulation of gene expression could not be supported with our experimental system.

REFERENCES

Van Den Abbeele, J., Claes, Y., van Bockstaele, D., Le Ray, D., and Coosemans, M. (1999). Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. Parasitology *118 (Pt 5)*, 469–478.

Agabian, N. (1990). Trans Splicing of Nuclear Pre-mRNAs. Cell *61*, 1157–1160.

Alirol, E., Schrumpf, D., Amici Heradi, J., Riedel, A., de Patoul, C., Quere, M., and Chappuis, F. (2013). Nifurtimox-eflornithine combination therapy for second-stage gambiense human african trypanosomiasis: medecins sans frontieres experience in the democratic republic of the congo. Clin. Infect. Dis. *56*, 195–203.

Alsford, S., Kawahara, T., Glover, L., and Horn, D. (2005). Tagging a T . brucei RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. Mol. Biochem. Parasitol. *144*, 142–148.

Anderson, S.A., Carter, V., and Parsons, M. (1998). Trypanosoma brucei: molecular cloning and stage-regulated expression of a malate dehydrogenase localized to the mitochondrion. Exp. Parasitol. *89*, 63–70.

Aranda, A., Maugeri, D., Uttaro, A.D., Opperdoes, F., Cazzulo, J.J., and Nowicki, C. (2006). The malate dehydrogenase isoforms from Trypanosoma brucei: subcellular localization and differential expression in bloodstream and procyclic forms. Int. J. Parasitol. *36*, 295–307.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. *38*, D457–D462.

Back, S.H., Schröder, M., Lee, K., Zhang, K., and Kaufman, R.J. (2005). ER stress signaling by regulated splicing: IRE1/HAC1/XBP1. Methods *35*, 395–416.

Bangs, J.D., Crain, P.F., Hashizume, T., McCloskey, J.A., and Boothroyd, J.C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. J. Biol. Chem. *267*, 9805–9815.

Berberof, M., Vanhamme, L., Tebabi, P., Pays, A., Jefferies, D., Welburn, S., and Pays, E. (1995). The 3'-terminal region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in Trypanosoma brucei. EMBO J. *14*, 2925–2934.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., et al. (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416–422.

Blattner, J., and Clayton, C. (1995). The 3′-untranslated regions from the Trypanosoma brucei phosphoglycerate kinase-encoding genes mediate developmental regulation. Gene *162*, 153–156.

Boothroyd, J.C., and Cross, G.A. (1982). Transcripts coding for variant surface glycoproteins of Trypanosoma brucei have a short, identical exon at their 5′ end. Gene *20*, 281–289.

Brecht, M., and Parsons, M. (1998). Changes in polysome profiles accompany trypanosome development. Mol. Biochem. Parasitol. *97*, 189–198.

Brenndörfer, M., and Boshart, M. (2010). Selection of reference genes for mRNA quantification in Trypanosoma brucei. Mol. Biochem. Parasitol. *172*, 52–55.

Bringaud, F., and Baltz, T. (1993). Differential Regulation of Two Distinct Families of Glucose Transporter Genes in Trypanosoma brucei. Mol. Cell. Biol. *13*, 1146–1154.

Capewell, P., Monk, S., Ivens, A., Macgregor, P., Fenn, K., Walrad, P., Bringaud, F., Smith, T.K., and Matthews, K.R. (2013). Regulation of Trypanosoma brucei Total and Polysomal mRNA during Development within Its Mammalian Host. PLoS One *8*, e67069.

Casero, R.A., Porter, C.W., and Bernacki, R.J. (1982). Activity of tunicamycin against Trypanosoma brucei in vitro and in vivo. Antimicrob. Agents Chemother. *22*, 1008–1011.

Clayton, C. (2002). Life without transcriptional control? From fly to man and back again. EMBO J. *21*, 1881–1888.

Clayton, C. (2013). The regulation of trypanosome gene expression by RNA-binding proteins. PLoS Pathog. *9*, e1003680.

Colasante, C., Robles, A., Li, C.-H., Schwede, A., Benz, C., Voncken, F., Guilbride, D.L., and Clayton, C. (2007). Regulated expression of glycosomal phosphoglycerate kinase in Trypanosoma brucei. Mol. Biochem. Parasitol. *151*, 193–204.

Cowling, V.H. (2010). Regulation of mRNA cap methylation. Biochem J *425*, 295–302.

Cowling, V.H., and Cole, M.D. (2010). Myc Regulation of mRNA Cap Methylation. Genes Cancer *1*, 576–579.

Cox, F.E.G. (2004). History of sleeping sickness (African trypanosomiasis). Infect. Dis. Clin. North Am. *18*, 231–245.

Cross, G.A. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology *71*, 393–417.

Cross, G.A. (1979). Crossreacting determinants in the C-terminal region of trypanosome variant surface antigens. Nature *277*, 310–312.

Daniels, J.-P., Gull, K., and Wickstead, B. (2010). Cell biology of the trypanosome genome. Microbiol. Mol. Biol. Rev. *74*, 552–569.

Denise, H., McNeil, K., Brooks, D.R., Alexander, J., Coombs, G.H., and Mottram, J.C. (2003). Expression of multiple CPB genes encoding cysteine proteases is required for Leishmania mexicana virulence in vivo. Infect. Immun. *71*, 3190–3195.

Dickson, L.M., and Brown, A.J. (1998). mRNA translation in yeast during entry into stationary phase. Mol. Gen. Genet. *259*, 282–293.

Djikeng, A., Shi, H., Tschudi, C., Shen, S., and Ullu, E. (2003). An siRNA ribonucleoprotein is found associated with polyribosomes in Trypanosoma brucei. RNA *9*, 802–808.

Doyle, J.J., Hirumi, H., Hirumi, K., Lupton, E.N., and Cross, G.A. (1980). Antigenic variation in clones of animal-infective Trypanosoma brucei derived and maintained in vitro. Parasitology *80*, 359–369.

Dutton, J. (1902). Note on a Trypanosoma occurring in the blood of man.

Ferguson, M., Homans, S., Dwek, R., and Rademacher, T. (1988). Glycosyl-phosphatidylinositol moiety that anchors Trypanosoma brucei variant surface glycoprotein to the membrane. Science*239*, 753–759.

Forde, R. (1902). Some clinical notes on a European patient in whose blood a Trypanosoma was observed.

Franco, J.R., Simarro, P.P., Diarra, A., Ruiz-Postigo, J.A., Samo, M., and Jannin, J.G. (2012). Monitoring the use of nifurtimox-eflornithine combination therapy (NECT) in the treatment of second stage gambiense human African trypanosomiasis. Res. Rep. Trop. Med. *3*, 93–101.

Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014). Epidemiology of human African trypanosomiasis. Clin. Epidemiol. *6*, 257–275.

Freire, E.R., Dhalia, R., Moura, D.M.N., da Costa Lima, T.D., Lima, R.P., Reis, C.R.S., Hughes, K., Figueiredo, R.C.B.Q., Standart, N., Carrington, M., et al. (2011). The four trypanosomatid eIF4E homologues fall into two separate groups, with distinct features in primary sequence and biological properties. Mol. Biochem. Parasitol. *176*, 25–36.

Freire, E.R., Malvezzi, A.M., Vashisht, A. a, Zuberek, J., Saada, E. a, Langousis, G., Nascimento, J.D.F., Moura, D., Darzynkiewicz, E., Hill, K., et al. (2014). Trypanosoma brucei translation initiation factor homolog EIF4E6 forms a tripartite cytosolic complex with EIF4G5 and a capping enzyme homolog. Eukaryot. Cell *13*, 896–908.

Freistadt, M., Cross, G.A., D.Branch, A., and D.Robertson, H. (1987). Direct analysis of the mini-exon donor RNA of Trypanosoma brucei: detection of a novel cap structure also present in messenger RNA. Nucleic Acids Res. *15*, 9861–9879.

Freistadt, M., Cross, G.A., and Robertson, H. (1988). Discontinuously synthesized mRNA from Trypanosoma brucei contains the highly methylated 5'cap structure, m7GpppA* A* C (2'-O) mU* A. J. Biol. Chem. *263*, 15071–15075.

Furuichi, Y., Morgan, M., Shatkin, A.J., Jelinek, W., Salditt-georgieff, M., and Darnell, J.E. (1975). Methylated, Blocked 5' Termini in HeLa Cell mRNA. Proc. Natl. Acad. Sci. U. S. A. *72*, 1904–1908.

Furuichi, Y., LaFiandra, A., and Shatkin, A.J. (1977). 5′-Terminal structure and mRNA stability. Nature *266*, 235–239.

Gilinger, G., and Bellofatto, V. (2001). Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic Acids Res. *29*, 1556–1564.

Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., McCarthy, J.E.G., and Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell *115*, 739–750.

Hall, M.P., and Ho, C.K. (2006a). Characterization of a Trypanosoma brucei RNA cap (guanine N-7) methyltransferase. RNA *12*, 488–497.

Hall, M.P., and Ho, C.K. (2006b). Functional characterization of a 48 kDa Trypanosoma brucei cap 2 RNA methyltransferase. Nucleic Acids Res. *34*, 5594–5602.

Hansen, K.D., Brenner, S.E., and Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. *38*, e131.

Hertz-Fowler, C., Ersfeld, K., and Gull, K. (2001). CAP5.5, a life-cycle-regulated, cytoskeleton-associated protein is a member of a novel family of calpain-related proteins in Trypanosoma brucei. Mol. Biochem. Parasitol. *116*, 25–34.

Hirumi, H., and Hirumi, K. (1994). Axenic culture of African trypanosome bloodstream forms. Parasitol. Today *10*, 80–84.

Ho, C.K., and Shuman, S. (2001). Trypanosoma brucei RNA triphosphatase. Antiprotozoal drug target and guide to eukaryotic phylogeny. J. Biol. Chem. *276*, 46182–46186.

Horn, D., and Cross, G.A. (1997). Analysis of Trypanosoma brucei vsg expression site switching in vitro. Mol. Biochem. Parasitol. *84*, 189–201.

Hotchkiss, T.L., Nerantzakis, G.E., Dills, S.C., Shang, L., and Read, L.K. (1999). Trypanosoma brucei poly(A) binding protein I cDNA cloning, expression, and binding to 5′ untranslated region sequence elements. Mol. Biochem. Parasitol. *98*, 117–129.

Huang, J., and van der Ploeg, L.H. (1991). Maturation of polycistronic pre-mRNA in Trypanosoma brucei: analysis of trans splicing and poly(A) addition at nascent RNA transcripts from the hsp70 locus. Mol. Cell. Biol. *11*, 3180–3190.

Huebner, P.W.A., and Milburn, R.M. (1980). Hydrolysis of pyrophosphate to orthophosphate promoted by cobalt(III). Evidence for the role of polynuclear species. Inorg. Chem. *19*, 1267–1272.

Imboden, M.A., Laird, P.W., Affolter, M., and Seebeck, T. (1987). Transcription of the intergenic regions of the tubulin gene cluster of Trypanosoma brucei: evidence for a polycistronic transcription unit in a eukaryote. Nucleic Acids Res. *15*, 7357–7368.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Ingolia, N.T., Brar, G. a, Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat. Protoc. *7*, 1534–1550.

Irmer, H., and Clayton, C. (2001). Degradation of the unstable EP1 mRNA in Trypanosoma brucei involves initial destruction of the 3′-untranslated region. Nucleic Acids Res. *29*, 4707–4715.

Jeffries, T.R., Morgan, G.W., and Field, M.C. (2002). TbRAB18, a developmentally regulated Golgi GTPase from Trypanosoma brucei. Mol. Biochem. Parasitol. *121*, 63–74.

Jensen, B.C., Sivam, D., Kifer, C.T., Myler, P.J., and Parsons, M. (2009). Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. BMC Genomics *10*, 482.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. Cell *51*, 273–281.

Kapotas, N., and Bellofatto, V. (1993). Differential response to RNA trans- splicing signals within the phosphoglycerate kinase gene cluster in Trypanosoma brucei. Nucleic Acids Res. *21*, 4067–4072.

Kolev, N.G., Franklin, J.B., Carmi, S., Shi, H., Michaeli, S., and Tschudi, C. (2010). The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. PLoS Pathog. *6*, e1001090.

Kolev, N.G., Ramey-Butler, K., Cross, G.A., Ullu, E., and Tschudi, C. (2012). Developmental progression to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. Science *338*, 1352–1353.

Laird, P.W., Zomerdijk, J.C., de Korte, D., and Borst, P. (1987). In vivo labelling of intermediates in the discontinuous synthesis of mRNAs in Trypanosoma brucei. EMBO J. *6*, 1055–1062.

Li, H., and Tschudi, C. (2005). Novel and essential subunits in the 300-kilodalton nuclear cap binding complex of Trypanosoma brucei. Mol. Cell. Biol. *25*, 2216–2226.

Liang, X., Haritan, A., Uliel, S., and Michaeli, S. (2003). trans and cis splicing in trypanosomatids: Mechanism, factors, and regulation. Eukaryot. Cell *2*, 830–840.

MacLean, L., Reiber, H., Kennedy, P.G.E., and Sternberg, J.M. (2012). Stage progression and neurological symptoms in Trypanosoma brucei rhodesiense sleeping sickness: role of the CNS inflammatory response. PLoS Negl. Trop. Dis. *6*, e1857.

Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H.O., Bishop, J.R., Falcone, F.H., Gavrilescu, C., Montgomery, J.L., Santori, M.I., et al. (2000a). A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. RNA *6*, 163–169.

Mair, G., Ullu, E., and Tschudi, C. (2000b). Cotranscriptional cap 4 formation on the Trypanosoma brucei spliced leader RNA. J. Biol. Chem. *275*, 28994–28999.

Martínez-Calvillo, S., Vizuet-de-Rueda, J.C., Florencio-Martínez, L.E., Manning-Cela, R.G., and Figueroa-Angulo, E.E. (2010). Gene expression in trypanosomatid parasites. J. Biomed. Biotechnol. 1–15.

Mašek, T., Valášek, L., and Pospíšek, M. (2011). Polysome analysis and RNA purification from sucrose gradients. In Methods in Molecular Biology (Clifton, N.J.), pp. 293–309.

Mazhari-Tabrizi, R., Eckert, V., Blank, M., Müller, R., Mumberg, D., Funk, M., and Schwarz, R.T. (1996). Cloning and functional expression of glycosyltransferases from parasitic protozoans by heterologous complementation in yeast: the dolichol phosphate mannose synthase from Trypanosoma brucei brucei. Biochem. J. *316 (Pt 3*, 853–858.

McNally, K.P., and Agabian, N. (1992). Trypanosoma brucei spliced-leader RNA methylations are required for trans splicing in vivo. Mol. Cell. Biol. *12*, 4844–4851.

Melville, S.E., Leech, V., Navarro, M., and Cross, G.A. (2000). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. Mol. Biochem. Parasitol. *111*, 261–273.

Milone, J., Wilusz, J., and Bellofatto, V. (2002). Identification of mRNA decapping activities and an ARE-regulated 3′ to 5′ exonuclease activity in trypanosome extracts. Nucleic Acids Res. *30*, 4040–4050.

Mittra, B., Zamudio, J.R., Bujnicki, J.M., Stepinski, J., Darzynkiewicz, E., Campbell, D. a, and Sturm, N.R. (2008). The TbMTr1 spliced leader RNA cap 1 2'-O-ribose methyltransferase from Trypanosoma brucei acts with substrate specificity. J. Biol. Chem. *283*, 3161–3172.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Mottram, J.C., Murphy, W.J., and Agabian, N. (1989). A transcriptional analysis of the Trypanosoma brucei hsp83 gene cluster. Mol. Biochem. Parasitol. *37*, 115–127.

Mottram, J.C., Coombs, G.H., and Alexander, J. (2004). Cysteine peptidases as virulence factors of Leishmania. Curr. Opin. Microbiol. *7*, 375–381.

Murphy, W.J., Watkins, K.P., and Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: Evidence for Trans splicing. Cell *47*, 517–525.

Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I., and Ochsenreiter, T. (2010). Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. PLoS Pathog. *6*, e1001037.

Obrig, T.G., Culp, W.J., Wallace, L., and Hardesty, B. (1971). The Mechanism by which Cycloheximide and Related Glutarimide Antibiotics Inhibit Peptide Synthesis on Reticulocyte Ribosomes.

Olego-Fernandez, S., Vaughan, S., Shaw, M.K., Gull, K., and Ginger, M.L. (2009). Cell morphogenesis of Trypanosoma brucei requires the paralogous, differentially expressed calpain-related proteins CAP5.5 and CAP5.5V. Protist *160*, 576–590.

Parsons, M., Nelson, R.G., Watkins, K.P., and Agabian, N. (1984). Trypanosome mRNAs share a common 5' spliced leader sequence. Cell *38*, 309–316.

Pays, E., and Nolan, D.P. (1998). Expression and function of surface proteins in Trypanosoma brucei. Mol. Biochem. Parasitol. *91*, 3–36.

Pereira, M.M.C., Malvezzi, A.M., Nascimento, L.M., Lima, T.D.C. da C., Alves, V.S., Palma, M.L., Freire, E.R., Moura, D.M.N., Reis, C.R.S., and de Melo Neto, O.P. (2013). The eIF4E subunits of two distinct trypanosomatid eIF4F complexes are subjected to differential post-translational modifications associated to distinct growth phases in culture. Mol. Biochem. Parasitol. *190*, 82–86.

Perry, K.L., Watkins, K.P., and Agabian, N. (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. U. S. A. *84*, 8190–8194.

Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H., and Boom, J.V. (1982). RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. Nucleic Acids Res. *10*, 3591–3604.

Pradet-Balade, B., Boulmé, F., Beug, H., Müllner, E.W., and Garcia-Sanz, J.A. (2001). Translation control: bridging the gap between genomics and proteomics? Trends Biochem. Sci. *26*, 225–229.

Preußer, C., Rossbach, O., Hung, L.-H., Li, D., and Bindereif, A. (2014). Genome-wide RNA-binding analysis of the trypanosome U1 snRNP proteins U1C and U1-70K reveals cis/trans-spliceosomal network. Nucleic Acids Res. *42*, 6603–6615.

Priotto, G., Kasparian, S., Mutombo, W., Ngouama, D., Ghorashian, S., Arnold, U., Ghabri, S., Baudin, E., Buard, V., Kazadi-Kyanza, S., et al. (2009). Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. Lancet *374*, 56–64.

Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J.D., and Clayton, C. (2009). Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC Genomics *10*, 495.

Roper, J.R., Guther, M.L.S., Milne, K.G., and Ferguson, M. a J. (2002). Galactose metabolism is essential for the African sleeping sickness parasite Trypanosoma brucei. Proc. Natl. Acad. Sci. U. S. A. *99*, 5884–5889.

Rosenthal, P.J. (1999). Advances in Parasitology Volume 43 (Elsevier).

Rottman, F., Shatkin, A.J., and Perry, R.P. (1974). Sequences containing methylated nucleotides at the 5′ termini of messenger RNAs: Possible implications for processing. Cell *3*, 197–199.

Ruan, J.-P., Shen, S., Ullu, E., and Tschudi, C. (2007). Evidence for a capping enzyme with specificity for the trypanosome spliced leader RNA. Mol. Biochem. Parasitol. *156*, 246–254.

Saas, J., Ziegelbauer, K., von Haeseler, A., Fast, B., and Boshart, M. (2000). A Developmentally Regulated Aconitase Related to Iron-regulatory Protein-1 Is Localized in the Cytoplasm and in the Mitochondrion of Trypanosoma brucei. J. Biol. Chem. *275*, 2745–2755.

Sagliocco, F.A., Vega Laso, M.R., Zhu, D., Tuite, M.F., McCarthy, J.E., and Brown, A.J. (1993). The influence of 5'-secondary structures upon ribosome binding to mRNA during translation in yeast. J. Biol. Chem. *268*, 26522–26530.

Sather, S., and Agabian, N. (1985). A 5' spliced leader is added in trans to both alpha- and beta-tubulin transcripts in Trypanosoma brucei. Proc. Natl. Acad. Sci. U. S. A. *82*, 5695–5699.

Sbicego, S., Vassella, E., Kurath, U., Blum, B., and Roditi, I. (1999). The use of transgenic Trypanosoma brucei to identify compounds inducing the differentiation of bloodstream forms to procyclic forms. Mol Biochem Parasitol *104*, 311–322.

Shatkin, A. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645–653.

Shuman, S. (2001). Structure, mechanism, and evolution of the mRNA capping apparatus. Prog. Nucleic Acid Res. Mol. Biol. *66*, 1–40.

Shuman, S. (2002). What messenger RNA capping tells us about eukaryotic evolution. Nat Rev Mol Cell Biol *3*, 619–625.

Siegel, T.N., Hekstra, D.R., Wang, X., Dewell, S., and Cross, G.A. (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. Nucleic Acids Res *38*, 4946–4957.

Silva, E., Ullu, E., Kobayashi, R., and Tschudi, C. (1998). Trypanosome capping enzymes display a novel two-domain structure. Mol. Cell. Biol. *18*, 4612–4619.

Simarro, P.P., Diarra, A., Ruiz Postigo, J.A., Franco, J.R., and Jannin, J.G. (2011). The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000-2009: the way forward. PLoS Negl. Trop. Dis. *5*, e1007.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A., Ruiz-Postigo, J.A., Fèvre, E.M., Mattioli, R.C., and Jannin, J.G. (2012). Estimating and mapping the population at risk of sleeping sickness. PLoS Negl. Trop. Dis. *6*, e1859.

Smith, T.K., Vasileva, N., Gluenz, E., Terry, S., Portman, N., Kramer, S., Carrington, M., Michaeli, S., Gull, K., and Rudenko, G. (2009). Blocking variant surface glycoprotein synthesis in Trypanosoma brucei triggers a general arrest in translation initiation. PLoS One *4*, e7532.

Sousa, M.C., Ferrero-Garcia, M.A., and Parodi, A.J. (1992). Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. Biochemistry *31*, 97–105.

Spirin, A.S. (1999). Ribosomes (Springer Science & Business Media).

Stephens, J.W.W., and Fantham, H.B. (1910). On the Peculiar Morphology of a Trypanosome from a Case of Sleeping Sickness and the Possibility of Its Being a New Species (T. rhodesiense). Proc. R. Soc. B Biol. Sci. *83*, 28–33.

Stöcklein, W., and Piepersberg, W. (1980). Binding of cycloheximide to ribosomes from wild-type and mutant strains of Saccharomyces cerevisiae. Antimicrob. Agents Chemother. *18*, 863–867.

Stoffel, S.A., Alibu, V.P., Hubert, J., Ebikeme, C., Portais, J.-C., Bringaud, F., Schweingruber, M.E., and Barrett, M.P. (2011). Transketolase in Trypanosoma brucei. Mol. Biochem. Parasitol. *179*, 1–7.

Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gürtler, R.E., McKerrow, J., Reed, S., and Tarleton, R. (2008). Kinetoplastids: related protozoan pathogens, different diseases. J. Clin. Invest. *118*, 1301–1310.

Sutton, R.E., and Boothroyd, J.C. (1986). Evidence for Trans splicing in trypanosomes. Cell *47*, 527–535.

Sutton, R.E., and Boothroyd, J.C. (1988). The Cap of Both Miniexon-Derived RNA and mRNA of Trypanosomes Is 7-Methylguanosine. Mol. Cell. Biol. *8*, 494–496.

Takagi, Y., Sindkar, S., Ekonomidis, D., Hall, M.P., and Ho, C.K. (2007). Trypanosoma brucei encodes a bifunctional capping enzyme essential for cap 4 formation on the spliced leader RNA. J. Biol. Chem. *282*, 15995–16005.

Tatu, U. (1997). Interactions between Newly Synthesized Glycoproteins, Calnexin and a Network of Resident Chaperones in the Endoplasmic Reticulum. J. Cell Biol. *136*, 555–565.

Trombetta, E.S., Simons, J.F., and Helenius, A. (1996). Endoplasmic Reticulum Glucosidase II Is Composed of a Catalytic Subunit, Conserved from Yeast to Mammals,

and a Tightly Bound Noncatalytic HDEL-containing Subunit. J. Biol. Chem. *271*, 27509–27516.

Tsai, B., Ye, Y., and Rapoport, T.A. (2002). Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. Nat. Rev. Mol. Cell Biol. *3*, 246–255.

Ullu, E., and Tschudi, C. (1991). Trans splicing in trypanosomes requires methylation of the 5' end of the spliced leader RNA. Proc. Natl. Acad. Sci. U. S. A. *88*, 10074–10078.

Ullu, E., and Tschudi, C. (1993). 2'-O-methyl RNA oligonucleotides identify two functional elements in the trypanosome spliced leader ribonucleoprotein particle. J. Biol. Chem. *268*, 13068–13073.

Ullu, E., and Tschudi, C. (1995). Accurate Modification of the Trypanosome Spliced Leader Cap Structure in a Homologous Cell-free System. J. Biol. Chem. *270*, 20365–20369.

Ullu, E., Matthews, K.R., and Tschudi, C. (1993). Temporal Order of RNA-Processing Reactions in Trypanosomes : Rapid trans Splicing Precedes Polyadenylation of Newly Synthesized Tubulin Transcripts. Mol. Cell. Biol. *13*, 720–725.

Urbaniak, M.D., Guther, M.L., and Ferguson, M.A. (2012). Comparative SILAC Proteomic Analysis of Trypanosoma brucei Bloodstream and Procyclic Lifecycle Stages. PLoS One *7*, e36619.

Urbaniak, M.D., Martin, D.M.A., and Ferguson, M.A.J. (2013). Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. J. Proteome Res. *12*, 2233–2244.

Vasquez, J.-J., Hon, C.-C., Vanselow, J.T., Schlosser, A., and Siegel, T.N. (2014). Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. Nucleic Acids Res. *42*, 3623–3637.

Vassilakos, A., Michalak, M., Lehrman, M.A., and Williams, D.B. (1998). Oligosaccharide binding characteristics of the molecular chaperones calnexin and calreticulin. Biochemistry *37*, 3480–3490.

Van Wazer, J.R., Griffith, E.J., and McCullough, J.F. (1955). Structure and Properties of the Condensed Phosphates. VII. Hydrolytic Degradation of Pyro- and Tripolyphosphate. J. Am. Chem. Soc. *77*, 287–291.

WHO (2014). WHO | Trypanosomiasis, human African (sleeping sickness).

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. Mol. Biochem. Parasitol. *99*, 89–101.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.

Yun, O., Priotto, G., Tong, J., Flevaud, L., and Chappuis, F. (2010). NECT is next: implementing the new drug combination therapy for Trypanosoma brucei gambiense sleeping sickness. PLoS Negl. Trop. Dis. *4*, e720.

Zinoviev, A., and Shapira, M. (2012). Evolutionary conservation and diversification of the translation initiation apparatus in trypanosomatids. Comp. Funct. Genomics 1–10.